PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 : C12Q 1/68, C07H 19/00

A1

(11) International Publication Number: WO 89/12694

(43) International Publication Date: 28 December 1989 (28.12.89)

(21) International Application Number: PCT/US89/02602

(22) International Filing Date: 20 June 1989 (20.06.89)

(30) Priority data:
209,247 20 June 1988 (20.06.88) US

(60) Parent Application or Grant
(63) Related by Continuation
US 209,247 (CIP)
Filed on 20 June 1988 (20.06.88)

(71) Applicant *(for all designated States except US)*: GENOMYX, INC. [US/US]; 453 Remillard Drive, Hillsboro, CA 94010 (US).

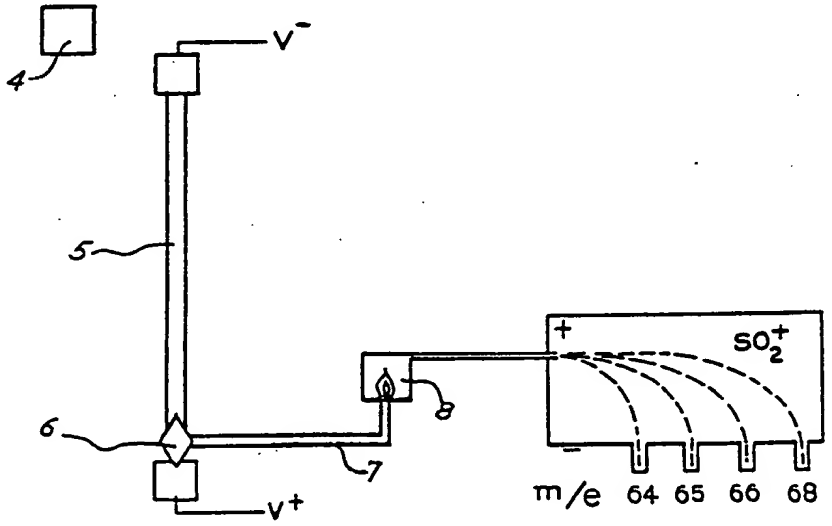
(72) Inventor; and
(75) Inventor/Applicant *(for US only)* : BRENNAN, Thomas, M. [US/US]; 453 Remillard Drive, Hillsboro, CA 94010 (US).

(74) Agents: McDONNELL, John, J. et al.; Allegretti & Witcoff, Ltd., 10 South Wacker Drive, Chicago, IL 60606 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent), US.

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: DETERMINING DNA SEQUENCES BY MASS SPECTROMETRY



(57) Abstract

This invention relates to the methods, apparatus, reagents and mixtures of reagents for sequencing natural or recombinant DNA and other polynucleotides. In particular, this invention relates to a method for sequencing polynucleotides based on mass spectrometry to determine which of the four bases (adenine, guanine, cytosine or thymine) is a component of the terminal nucleotide. In particular, the present invention relates to identifying the individual nucleotides by the mass of stable nuclide markers contained within either the dideoxynucleotides, the DNA primer, or the deoxynucleotide added to the primer. This invention is particularly useful in identifying specific DNA sequences in very small quantities in biological products produced by fermentation or other genetic engineering techniques. The invention is therefore useful in evaluating safety and ther health concerns related to th presence of DNA in products resulting from genetic engineering techniques.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	ML	Mali
AU	Australia	FR	France	MR	Mauritania
BB	Barbados	GA	Gabon	MW	Malawi
BE	Belgium	GB	United Kingdom	NL	Netherlands
BF	Burkina Fasso	HU	Hungary	NO	Norway
BG	Bulgaria	IT	Italy	RO	Romania
BJ	Benin	JP	Japan	SD	Sudan
BR	Brazil	KP	Democratic People's Republic of Korea	SE	Sweden
CF	Central African Republic			SN	Senegal
CG	Congo	KR	Republic of Korea	SU	Soviet Union
CH	Switzerland	LI	Liechtenstein	TD	Chad
CM	Cameroon	LK	Sri Lanka	TG	Togo
DE	Germany, Federal Republic of	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		
ES	Spain	MG	Madagascar		

DETERMINING DNA SEQUENCES BY MASS SPECTROMETRY

BACKGROUND OF THE INVENTION

A. Field of the Invention

This invention relates to the field of the determination of DNA sequences and the uses of automated techniques for such determination.

The ability to sequence DNA has become a core technology in molecular biology, and has contributed greatly to the understanding of DNA structural organization and gene function. The facility with which DNA sequencing may be accomplished will substantially affect the rate of development of related technologies, including the production of new therapeutic agents, useful plant varieties and microorganisms via recombinant DNA technology and the understanding of human genetic disorders and pathology through gene mapping and chromosomal sequence analysis.

Initially, researchers focused on reading the genetic code and the translation of the nucleotide sequence into the amino acid sequence of a protein. This occurs by a process of DNA transcription into mRNA, and then actual synthesis of the protein on ribosomes. In eucaryotic cells, large specific segments of the initial transcript of mRNA, termed introns, are transcribed but are excised during an intermediary processing step. Much of the chromosomal DNA is not translated, and its specific function is largely unknown. This "intervening" or intron DNA was first thought to be excess genetic material. However, as biologists begin to unravel the details of cell differentiation and the processes controlling gene transcription it is now believed that the specific sequences of certain portions of some of these large regions of transcribed but untranslated DNA may also provide important regulatory signals.

The potential applications which derive from DNA sequencing have only begun to be explored. On large scale, analysis of human chromosomal DNA is considered vital to understanding human pathological conditions, including genetic diseases, AIDS and cancer, because often only subtle differences, even single nucleotide substitutions, can lead to serious dis rders. Serious consideration is now being given to the sequencing of the entire human genome - approximately 3 billion base pairs. The success of this project will depend on rapid, sensitive, inexpensive automated methods to sequence DNA.

The fundamental approach to determination of DNA sequence has been well established. Restriction endonucleases are employed to cleave chromosomal DNA into specific smaller segments, and recombinant cloning techniques are then used to purify and generate analyzable quantities of DNA. The specific sequence of each segment can then be determined by either the Maxam-Gilbert chemical cleavage, or preferably, the Sanger dideoxy terminated enzymatic method. In either case, a set of all possible fragments ending in a specific base are generated. The individual fragments can be resolved electrophoretically by molecular weight, and the sequence on the original DNA segment is then derived by knowing the identity of the terminal base in each fragment.

In its broadest aspect, this invention is directed to methods and reagents for sequencing DNA and other polynucleotides. In particular, this invention describes reagents and methods for automating and increasing the sensitivity of both the Sanger, Proc. Natl. Acad. Sci. USA, 74, 5463 (1977) and Gish and Eckstein, Science, 240, 1520-1522 (1988), procedures for sequencing polynucleotides. The methods of the present invention are based on mass spectrometric determination of each of the four component terminal nucleotide residues, where the information regarding the identity of the individual nucleotides is contained in the mass of stable nuclide markers.

B. Summary Of The Prior Art

In the Sanger dideoxy method (Proc. Natl. Acad. Sci. USA, 74, 5463 (1977)), the DNA to be sequenced is exposed to a DNA polymerase, a cDNA primer, and a mixture of the four component deoxynucleotides, plus one of the four possible 2,3-dideoxy nucleotides. The DNA to be sequenced is typically a single stranded DNA clone prepared in the phage vector M13, although Chen and Seeburg have disclosed a method for applying the Sanger method to supercoiled plasmid DNA (DNA 4:165-170 (1985)). In addition, Innis et al., Proc. Natl. Acad. Sci., USA 85, 9436-9440 (1988) have disclosed a method for direct sequencing of chromosomal DNA amplified by the polymerase chain reaction. For any DNA template, however, the principle behind the dideoxy chain termination method remains the same. There is a competition for incorporation of the normal deoxy- and the dideoxy-nucleotide by the polymerase into the growing complementary chain. When a dideoxy nucleotide is incorporated, further chain xtension is prevented. Since there is a finite probability that this chain terminating event may

occur at each complementary site of the appropriate base, a mixture of all possible fragments ending in that dideoxy base will be generated. This mixture of fragments can be separated by size via gel electrophoresis. When the experiment is repeated with each dideoxy base, four mixtures of fragments, each terminating in a specific residue are produced. When this set of mixtures is chromatographed in four adjacent lanes, so that fragment lengths in the four mixtures can be correlated with each other, the sequence of the original DNA is determined by relating the fragment length to the identity of the terminating dideoxy base.

Maxam and Gilbert, Methods in Enzymology, 65, 499-500 (1980), disclosed a method for DNA sequencing using chemical cleavage. In this method, each end of a DNA fragment to be sequenced is labeled. This DNA fragment is then cleaved preferentially at one of the nucleotides, under conditions favoring one cleavage per strand. This procedure is then repeated for each of the other three nucleotides. The four samples are then run side by side on an electrophoretic gel. Autoradiography identifies the position of a particular nucleotide by the length of the fragments produced by cleavage at that particular nucleotide. This method suffers from the same drawbacks as the Sanger method.

The position of the fragment in gel electrophoresis is usually revealed by staining or by autoradiography. In autoradiography methods, the fragments have typically been labeled with ^{32}P or ^{35}S radionuclides where either the DNA primer or one of the component deoxynucleotides have been tagged, and that label incorporated in a specific or random fashion. After fractionation of the fragment on acrylamide gels, the gels are used to expose films. This presents a number of difficulties. For example, the short half-life of ^{32}P requires that the sequencing experiment be anticipated days in advance so that fresh label can be used. Additionally, the high energy beta radiation emitted by the ^{32}P leads to cission of the phosphodiester linkages within the DNA fragments synthesized in the sequencing reaction and thus requires immediate fractionation of sequencing reaction products. The use of ^{35}S (Ornstein, et al., Biotechniques, 3, 476 (1985), which has a longer half-life and less energetic emission somewhat ameliorates these problems, but requires much longer times of exposure to film for the development of a usable autoradiograph, often in the range of one to three days. Whichever radionuclide is used, the fact that a single

type of label is used for each sequencing reaction requires that each set of reaction products be fractionated in a separate lane on the sequencing gel. Common problems in running sequencing gels include uneven heating and the presence of impurities, either of which can cause adjacent lanes on the sequencing gel to run in an uneven fashion making the comparison of fragment migration in adjacent lanes, and thus DNA sequence determination, difficult or impossible. The use of unstable radionuclides also poses a health risk to the investigator.

An alternate method of detection was developed by the California Institute of Technology group (Smith, et al., Nature, 321, 674 (1986)) in which the terminal base residues are labeled with a fluorescent marker attached to the DNA primer. In four fluorescent markers of different spectral emission maxima are used, then the four separate sets of polymerase fragments can be combined and co-chromatographed. This method is also disclosed in EPO Patent No. 87300998.9.

A second variation of the fluorescent tagging approach has recently been reported by the DuPont group (Science, 238, 336 (1987)) wherein a unique fluorescent moiety is attached directly to the dideoxy nucleotide. This may represent an improvement over the CalTech primer tagging approach in that a single polymerase experiment can now be run with a mixture of the four dideoxy terminating bases. However, one trade-off for this simplification is potential replication errors by the polymerase, arising from mis-incorporation of the modified dideoxynucleotide base analogs.

These modified Sanger methods are an improvement over the original Sanger method in the extent to which DNA can be sequenced because the chromatographic ambiguities have been reduced. However, a number of limitations are associated with the use of fluorescent labels in these modified Sanger reactions. In particular, there are chromatographic differences among fragments arising from the unique mobilities of the different organic fluorescent markers. Moreover, there are difficulties in distinguishing individual fluorescent markers because of overlap in their spectral bandwidths. Finally, there is a low sensitivity of detection inherent in the extinction coefficients of the fluorescent markers.

All of the above variants of the Sanger method for sequencing have used slab gel electrophoresis to effect size separation of the DNA fragments. The casting and loading of slab gels is a skilled but

intrinsically manual operation. The only aspect of this process which has been automated with any success is the reading of the gel by certain commercial devices with some type of laser scanner/spectrophotometer.

A labeling method is needed which eliminates chromatographic ambiguity by imparting to each sequencing reaction product its own specific tag, but in which this specific tag is "invisible" to the chromatographic apparatus, i.e., does not affect the chromatographic mobility of the different sequencing products differentially. Additionally, a label detection system is needed which is much more sensitive than the fluorescence system, and which can make distinction in labels based upon characteristics which separate them discretely, rather than by trying to distinguish between broad overlapping traits. Ideally, a stable, non-radioactive label would be used eliminating the short useful lifetime of the label and products containing the label, as well as potential health risks to investigators.

Eckstein and Goody, Biochemistry, 15, 1685 (1976), discloses a method of chemical synthesis for adenosine-5'-(O-1-thiotriphosphate) and adenosine-5'-(O-2-thiotriphosphate).

Eckstein, Accounts Chem. Res., 12, 204 (1978), discloses a group of phosphorothioate analogs of nucleotides.

Gish and Eckstein, Science, 240, 1520-1522 (1988), disclose an alternative method for sequencing DNA and RNA employing base specific chemical cleavage of phosphothioate analogs of the nucleotides which were incorporated in a cDNA sequence.

Japanese Patent No. 59-131,909 (1986), discloses a nucleic acid detection apparatus which detects nucleic acid fragments which are separated by electrophoretic techniques, liquid chromatography, or high speed gel filtration. Detection is achieved by utilizing nucleic acids into which S, Br, I, or Ag, Au, Pt, Os, Hg or similar metallic elements have been introduced. These elements are generally absent in natural nucleic acids. Introduction of one of these elements into a nucleotide of a nucleic acid allows that nucleic acid or fragment thereof to be detected by means of atomic absorption, plasma emission or mass spectroscopy. However, this reference does not suggest or disclose any application of the described methods or apparatus to the sequencing of DNA, such as by the Sanger method. Specifically, it does not teach that a plurality of specific isotopes

may be used to identify the specific terminal nucleotide residues. Nor does it teach that by total combustion of DNA to oxides of carbon, hydrogen, nitrogen and phosphorus, the detection sensitivity by mass spectrometry for trace elements, such as sulfur which is not normally found in DNA, is vastly improved. The combustion step, which is one aspect of the present application, is essential to eliminate the myriad of fragment ions from DNA. These fragment ions would normally mask the presence of trace ions of SO_2 in conventional mass spectrometry. What this reference does disclose is that DNA may be tagged (by undisclosed means) with trace elements, including sulfur, as an aid to detection of DNA, and that these trace elements may be detected by a variety of means, including mass spectrometry.

Details of DNA sequencing are found in Current Protocol In Molecular Biology, John Wiley & Son, N.Y., N.Y., F. M. Ansubel, et al., eds., (1987), Chapter 7 of which is hereby incorporated by reference. Smith, et al., Anal. Chem., 60, 438-441 (1988), describes capillary zone electrophoresis-mass spectrometry using an electrospray ionization interface and is hereby incorporated by reference.

SUMMARY OF THE INVENTION

This invention relates to improved methods for sequencing DNA, DNA fragments, or other polynucleotides. The invention includes apparatus, reagents and mixtures of reagents for carrying out the method. In particular, this invention relates to the use of mass spectrometry to identify the terminal nucleotide of a polynucleotide, based upon the presence of a specific stable nuclide marker in the terminal nucleotide or the polynucleotide fragment containing that particular terminal nucleotide. The invention offers numerous advantages over previous methods of sequencing polynucleotides, including greater sensitivity, increased signal specificity, simplified manipulation and safer handling.

BRIEF DESCRIPTION OF THE FIGURES

FIGURE 1 shows a schematic diagram of a complementary DNA sequence attached to a primer DNA sequence and a typical series of chain terminated polynucleotide fragments prepared according to Scheme E herein.

FIGURE 2A shows in combination a column for separating DNA sequences according to size and a means for sequentially transporting DNA sequences to a mass spectrometer.

FIGURE 2B shows superimposed "ion current vs. time" printouts for $^{32}SO_2$, $^{33}SO_2$, $^{34}SO_2$, and $^{32}SO_2$, resulting from combustion of a chain terminated DNA sequence.

BRIEF DESCRIPTION OF THE INVENTION

This invention relates to methods, reagents, apparatus and intermediates involved in the determination of natural or artificially made ("recombinant") DNA sequences and fragments thereof. This invention seeks to eliminate numerous deficiencies in the prior art by embodying greater convenience, less chromatographic ambiguity, greater sensitivity and safer handling than existing procedures. In particular, this invention involves the determination of DNA sequences using a combination of chain termination DNA sequencing techniques and mass spectroscopy. Thus, in a typical chain terminating DNA sequencing determination such as taught by Sanger, et al., Proc. Natl. Acad. Sci. USA, 74, 5463, (1977) involving a DNA primer, deoxynucleotidetriphosphates, dideoxynucleotidetriphosphates in the presence of a DNA polymerase, such as Klenow fragment, are used to determine the DNA sequence. However, in embodiments of the present invention the DNA primer, the deoxynucleotides or the dideoxynucleotides are labeled with isotopes detectable by mass spectrometry to determine the DNA sequence. For example, if the dideoxynucleotides (A, G, C, T) triphosphates, abbreviated as ddATP, ddGTP, ddCTP and ddTTP respectively, are labeled with isotopes of different masses respectively, and chain terminated fragments corresponding to those fragments are separated and analyzed by mass spectrometry, a direct reading of the DNA sequence is obtained. Generally, the labeled component of each dideoxynucleotide component of the chain terminated DNA sequence is converted to a more convenient species for mass spectrometry determination, i.e. sulfur isotopes are oxidized to sulfur dioxide. If the DNA primer or deoxynucleotides are labeled, reactions between specifically labeled deoxynucleotides must be first carried out in the presence of a specific dideoxynucleotide. This is necessary so that a specific label is associated with a specific chain terminated DNA sequence. Once the individual reactions are conducted, the chain terminated DNA sequences can be mixed, separated, and analyzed by mass spectrometry because there will then be a specific relationship between a specific isotope and the terminal dideoxynucleotide. This invention is much more sensitive than existing systems and therefore is especially useful in

determining the sequence of small quantities of DNA which are contaminants in products resulting from fermentation and other biotechnology related processes, i.e. for "screening" applications. The invention also includes reagents and analytical instruments for carrying out the above methods as well as intermediate mixtures of chain terminated DNA sequences produced while carrying out the methods of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

This invention relates to an improved method for sequencing polynucleotides using mass spectrometry to determine which of the four bases (adenine, guanine, cytosine or thymine) is a component of the terminal nucleotide. In particular, the present invention relates to identifying the individual nucleotides in a DNA sequence by the mass of stable nuclide markers contained within either the dideoxynucleotides, the DNA primer, or the deoxynucleotides added to the primer. The invention also includes reagents and analytical instruments for carrying out the above methods as well as mixtures of chain terminated DNA sequences.

Information regarding the identity of the terminal base in a particular fragment may be signified by using a unique isotopic label for each of the four bases. The determination of which isotope marker is present, and thus which terminal base a fragment contains, can then be readily accomplished by mass spectral methods. Detection of ions by mass spectra is perhaps the most sensitive physical method available to the analytical chemist, and represents orders of magnitude better sensitivity than optical detection of fluorescence.

If stable isotopes are chosen for labeling, then the isotope ratios are fixed by the mode of synthesis. The group of suitable atomic ("nuclide") markers include those from carbon (^{12}C / ^{13}C), chlorine (^{35}Cl / ^{37}Cl), bromine (^{79}Br / ^{81}Br) and sulphur (^{32}S / ^{33}S / ^{34}S / ^{36}S). Since sulfur, chlorine and bromine are not normal constituents of DNA, i.e. they are "foreign", analysis for those foreign isotopes does not require consideration of their natural abundance ratio. It is noted that sulfur is unique among this group in that it alone contains four stable isotopes, each of which can be used to represent one of the four nucleotide bases.

Further, if the fragments are subjected to combustion, then a light volatile derivative of the marker atom can be detected. Combustion

converts DNA to the oxides of carbon, hydrogen, nitrogen and phosphorus. The inclusion of a combustion step enormously simplifies the detection of trace atoms because it eliminates the problem of producing and analyzing high mass molecular ions.

With sulfur, combustion of the polynucleotide fragments in a hydrogen-oxygen flame or pyrolysis tube will yield sulfur dioxide (SO_2). Thus, the terminal base of the fragment may be identified by determining the mass of the SO_2 ion as 64, 65, 66, or 68. This is a simple distinction by existing mass spectral devices using either quadrupole or permanent magnet analyzers. For a permanent magnet device, a set of four permanently fixed ion detectors can be mounted to continuously monitor the individual ion currents. A quadrupole analyzer with a single ion-multiplier detector is presently preferred.

There are numerous ways in which a marker isotope could be incorporated into the complementary DNA fragment. These include substituting the marker isotope on the pyrimidine, purine, or ribose moieties, or the phosphate bridges between individual nucleotides. Further, the marker isotope may be contained in part of the cDNA primer, randomly incorporated along the chain in one or several of the deoxy-base units, or specifically in the terminal dideoxy residue. The only restriction is that the particular substitution be unique for that particular set of fragments.

The site for the stable sulfur label is most preferably the phosphate bridge, using labeled thiophosphate in place of ordinary phosphate. The technique for inserting a stable thiophosphate label in place of ordinary phosphate is similar to that employed in conventional ^{35}S radiolabeling experiments. The chemistry and enzymology of the polymerase reaction using deoxynucleotide-α-thiotriphosphates have been investigated extensively. Any future developments in cloning vectors or polymerase enzymes should also be able to utilize the thiophosphate derivatives of the present invention.

If the isotope label is to be incorporated into the cDNA primer or randomly along the chain as a deoxy-base surrogate, then it is necessary to perform a separate polymerase experiment with each of the appropriate dideoxy-base residues prior to mixing and chromatography. The advantage of using primer or intra-chain labeling is that several atoms of the marker

isotope may be incorporated per mole of DNA fragment, and thus enhance detection sensitivity.

If, on the other hand, the isotope label is contained in the dideoxy base itself, then it is not necessary to perform individual polymerase experiments. Instead, a mixture of the four dideoxy bases, each with a unique isotope label, together with a mixture of the four normal deoxy bases in stoichiometric ratios appropriate for the specific polymerase enzyme could be used to generate the complete set of labeled fragments in a single polymerase experiment. Each fragment, regardless of its size, will contain one atom of the marker isotope on its terminal (dideoxy) nucleotide, wherein the marker isotope would indicate the identity of the terminal nucleotide.

Schemes A and B below illustrate typical sulfur and halogen labeling respectively,

wherein, an asterisk (*) is used herein to indicate the presence of an isotopic label in accordance with the present invention;

wherein by A, T, G, and C is meant the bases adenine, thymine, guanine, and cytosine respectively;

wherein by "Alk" is meant straight or branched chain lower alkyl of 1-6 carbon atoms;

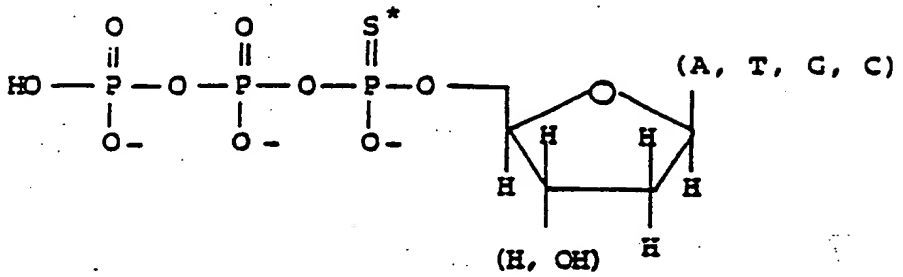
wherein by "S*" is meant a sulfur isotope of the group consisting of ^{32}S, ^{33}S, ^{34}S and ^{36}S with the proviso that each isotope be uniquely associated with a member of the group consisting of A, T, G, and C respectively; and

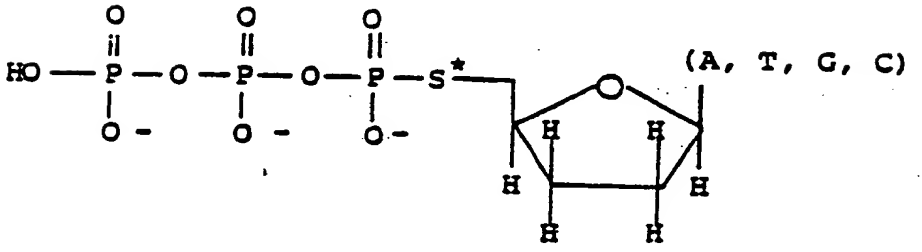
wherein by "X*" is meant a "halogen" isotope of the group consisting of ^{35}Cl, ^{37}Cl, ^{79}Br and ^{81}Br with the proviso that each isotope be uniquely associated with a member of the group consisting of A, T, G and C respectively.

Scheme A

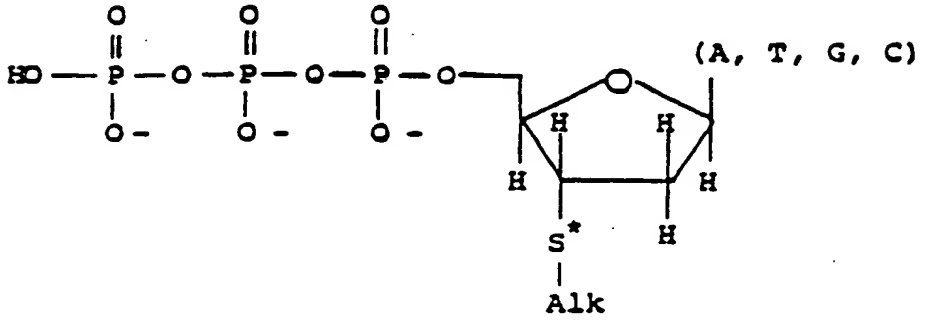
Sulfur Labeling

(a) at the phosphate bridge of "Rib":

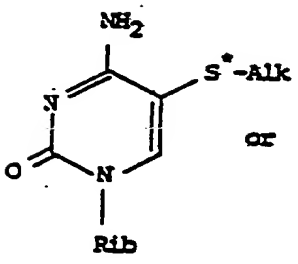




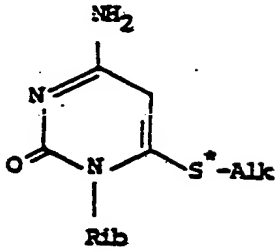
(b) on the deoxyribose:



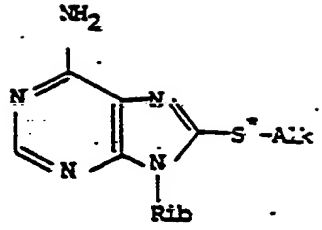
(c) on the base component:



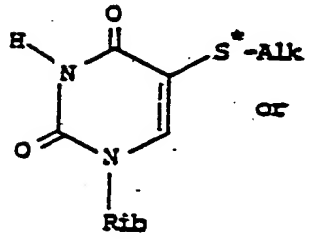
or



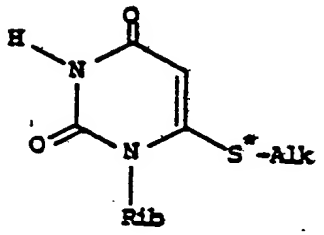
C*



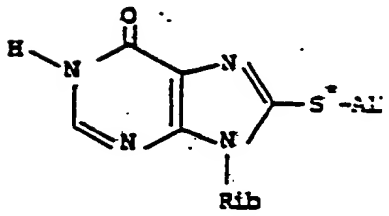
A*



or



T*

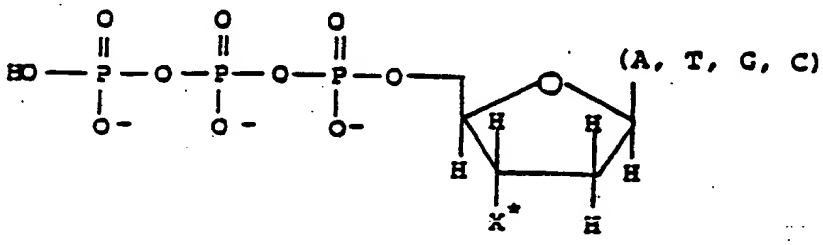


G*

SCHEME B

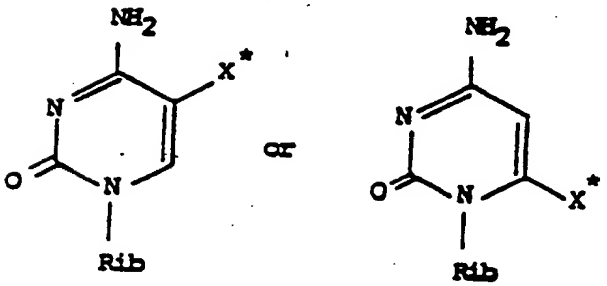
Chlorine/Bromine Labeling

(a) On the deoxyribose

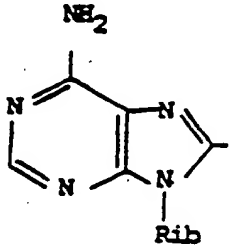


$X^* = Cl^*$ or Br^*

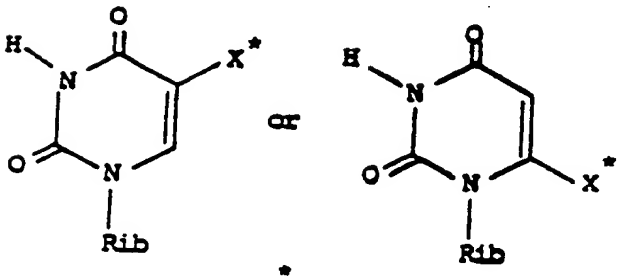
(b) On the base component



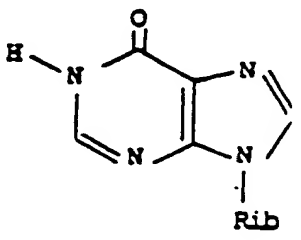
C^*



A^*



T^*



G^*

Labeling Schemes C, D, and E below show three ways in which specific isotopes, designated as *1, *2, *3, and *4 can be uniquely associated with specific terminal nucleotides in a terminated complementary DNA (cDNA) sequence. For convenience in Schemes C-E, the "TP" designation for the deoxy and dideoxy triphosphates has been deleted. In Schemes C and D, the dideoxy chain terminating reaction is conducted separately and then the terminated chains are mixed prior to separation. In particular, Scheme D is a modification of the chemical cleavage procedure of Gish and Eckstein, Science, 240, 1520 (1988) whereby the DNA fragment undergoes selective alkaline cleavage adjacent to the phosphothioate linkage, leaving a labeled deoxy compound as the terminal nucleotide in the fragment. In Scheme D, one creates a series of such fragments which differ from one another solely in size, via the presence of an additional terminal nucleotide. Identification of each terminal nucleotide (via each isotopic marker) in relation to size of the fragment provides the base sequence of the DNA or polynucleotide of interest.

In Scheme E, a mixture of the four individually labeled dideoxynucleotide triphosphates, together with a mixture of the four deoxy nucleotide triphosphates are reacted together with the primer ("P") in a single reaction. Because only one reaction and one separation need be run in Scheme E, it can be readily seen that the labeled dideoxy scheme, Scheme E, is the preferred method of the present invention. Particularly preferred reactants in Scheme E are the labeled dd(A*, C*, G*, or T*) triphosphates where the labels ^{32}S, ^{33}S, ^{34}S and ^{36}S replace a phosphate oxygen as shown in Scheme A.

SCHEME C

Labeled Primers

1. $P_{A^{*1}}$ + d(A,C,G.T) + dd(A) => $P_{A^{*1}}$dN_dN. . . . .dd(A)
2. $P_{C^{*2}}$ + " + dd(C) => $P_{C^{*2}}$dN_dN. . . . .dd(C)
3. $P_{G^{*3}}$ + " + dd(G) => $P_{G^{*3}}$dN_dN. . . . .dd(G)
4. $P_{T^{*4}}$ + " + dd(T) => $P_{T^{*4}}$dN_dN. . . . .dd(T)

SCHEME D

Labeled Deoxy

1. P + d(A,C,G,T) + d(A^{*1}) + dd(A) => P. .dN_dA^{*1}dN. . .dd(A)
2. " + " + d(C^{*2}) + dd(C) => P. .dN_dC^{*2}dN. . .dd(C)
3. " + " + d(G^{*3}) + dd(G) => P. .dN_dG^{*3}dN. . .dd(G)
4. " + " + d(T^{*4}) = dd(T) => P. .dN_dT^{*4}dN. . .dd(T)

SCHEME E

Labeled Dideoxy

1. P + d(A,C,G,T) + dd(A^{*1},C^{*2},G^{*3},T^{*4}) => P. .dN_dN. . . . . dd(A^{*1})
   P. .dN_dN. . . . . dd(C^{*2})
   P. .dN_dN. . . . . dd(G^{*3})
   P. .dN_dN. . . . . dd(T^{*4})

Accordingly, the invention not only includes reagents but also a mixture of unique isotopically labeled dideoxynucleotide triphosphates (ddC*TP, ddG*TP, ddT*TP and ddA*TP) where each dideoxynucleotide triphosphate is labeled with a different sulfur or halogen isotope. In particular, sulfur labeling, consisting of the isotopes ^{32}S, ^{33}S, ^{34}S and ^{36}S, is preferred.

The invention also includes the intermediate mixture of dideoxy chain-terminated DNA sequences in which each chain-terminated DNA sequence contains an isotope measurable by mass spectrometry and in which each isotope relates to a specific chain-terminating dideoxynucleotide. The stable nuclide marker may be incorporated into the DNA primer, the DNA chain extended from the primer, or the dideoxynucleotide which terminates the DNA chain extended from the primer. The invention also includes the mixture of isotope-labeled, chain-terminated DNA sequences separated by size.

Although this invention has been discussed in terms of sequencing DNA, the method and the reagents of the present invention could also be used to sequence RNA by providing reverse transcryptase as the polymerase.

Figure 1 illustrates a complementary DNA sequence attached to a promoter DNA sequence and typical series of chain terminated polynucleotide fragments prepared according to Scheme E from mixtures of deoxynucleotide triphosphates and labeled dideoxynucleotide triphosphates. These labeled fragments illustrate labeled chain terminated complementary DNA sequences 1 prepared by the method of the present invention, wherein the size of each complementary DNA fragment corresponds to the relative position of that fragment's terminal nucleotide in the overall complementary DNA sequence. These labeled fragments sequences are separated by size by an electrophoresis column 2. The fragments from the electrophoresis column 2 are sequentially eluted to a detector 3. Figure 2A shows in more detail the apparatus for determining a DNA sequence. DNA sequences are prepared in reaction chamber 4. The mixture of labeled terminated DNA fragments are separated according to size by electrophoresis on a polyacrylamide gel column 5 wherein migration occurs from the cathode (V^-) to the anode (V^+). The fractions are taken off the polyacrylamide gel column 5 sequentially by size at transfer point 6 where is provided a means

7 for transferring the terminated DNA fragments to an oxidizer or combustion chamber 8. In the oxidizer or combustion chamber 8, the sulfur label is oxidized to SO_2 and the labeled SO_2 is detected in a mass spectrometer 9. Figure 2B shows typical superimposed "ion current v. time" plots for m/e 64, 65, 66, and 68, corresponding to ions produced by the four stable isotopes of sulfur, i.e., $^{32}SO_2$, $^{33}SO_2$, $^{34}SO_2$ and $^{36}SO_2$, respectively. When the stable isotopes of sulfur associated with the bases, A, C, G and T are ^{32}S, ^{33}S, ^{34}S and ^{36}S respectively, a plot corresponding to the DNA sequence illustrated at the top of Figure 2B is obtained. In this manner, the DNA sequence of any genetic material can be determined automatically and on femto- to nanomolar quantities of material.

There are several variations in the design of an automated DNA sequencer of the present invention. The major components of the device are the reaction chamber for conducting polymerase reactions, the chromatographic device consisting of some form of electrophoresis, the effluent transport, the combustion system, and the mass spectral analyzer. Because this instrument is designed to operate on femto- to nano-molar quantities of DNA, it is important that the geometry of all component systems be kept to a minimum size.

The chromatographic system may be of a laned plate or tubular configuration. In the plate designs, the supporting medium for the chromatographic separation will be most preferably a polyacrylamide gel, where the ratio of acrylamide to bis-acrylamide is more preferably between 10:1 and 100:1. Although persulfate is the typical polymerization catalyst used by most workers to prepare polyacrylamide gel plates, the background of sulfate ions may be unacceptably high without extensive washing. Ultraviolet irradiation can be used successfully to initiate cross-linking and produce high quality gels, which can be used immediately without washing.

For tubular designs, the chromatographic separation may be conducted in a gel-filled capillary or in an open tubular configuration. The preferred dimensions of the capillary depend on whether an open or gel-filled medium is selected. For gel-filled devices, preferred diameters are 50 to 300 microns. In open tubular configurations, however, the preferred diameters are 1 to 50 microns. The preferred length of the capillary depends on the diameter and the amount of DNA sample which will be applied, as well as

the field strength of the applied electrophoretic voltage. The preferred length is optimally between 0.25 and 5 meters.

For open tubular configurations, the capillary will preferably be fabricated from fused silica. Under typical operating conditions, where pH of the buffer is usually maintained in a range of 5.0 to 11.0, the surface of the silica will have a net negative charge. This surface charge establishes conditions in which there is a bulk electrosmotic flow of buffer toward the negative electrode. The DNA fragments also possess a negative charge and are attracted to the positive electrode. Hence, they will therefore move more slowly than the bulk electroosmotic flow. In gel-filled devices, the supporting medium minimizes electroosmosis. Since the gel has no charge, the negatively charged DNA fragments migrate toward the positive electrode.

There are a variety of techniques to modify the surface charge on the wall of an open capillary. One particularly useful method is to covalently modify the wall with a monomer such as methacryloxypropyltrimethoxysilane. This monomer can then be crosslinked with the acrylamide to give a thin bonded monolayer which is similar in characteristics to the polyacrylamide gel-filled capillaries. The distinct advantage of the coated wall method is that the capillary can be recycled after each analysis run simply by flushing with fresh buffer. (See S. Hjerton, J. Chromatography, 347, 191, (1985) for details).

The transfer system is selected to match the particular chromatographic design. The chromatographic system may be of laned plate or tubular configuration. It is desirable to have the chromatography effluent in a closed environment. The tubular configurations may be more amenable to sample transfer designs which pump, spray or aspirate the column effluate into the combustion chamber, and thus minimize degradation in resolution because of post-chromatographic remixing of fractions.

For the open plate type devices, a moving belt or wire system can be used satisfactorily. In this system, a thin coating of the column effluent is spread on the ribbon to transport the eluted fractions through a pre-drying oven and then into the combustion furnace. The transport ribbon may be fabricated from platinum or other noble metal, and may be continuously looped because the ribbon can be effectively cleaned upon passage through the combustion furnace. Less preferably, the ribbon may

be fabricated from a glass or ceramic fiber or carbon steel. In this design, the ribbon would be taken up on a drum for disposal.

For tubular configurations, the most preferred embodiment of the transfer system is to use an electrospray nebulizer method to create a fine aerosol of the column effluent. In this technique, a small charge of optimally less than 3000 volts is applied to the emerging droplet. The charge of a larger droplet tends to disperse it into a very fine mist of singly charged droplets. These fine charged droplets can be focused and directed by electric or magnetic fields, much as in an ink-jet printer head. It is important to control the temperature of the flowing gas stream into which the aerosol is introduced. If the temperature is too great, then the droplets will tend to evaporate on the capillary injector. If the temperature is too low to overcome the surface tension of the effluent, the individual droplets will not be adequately dispersed. The composition and ionic strength of the supporting electrolyte is important. The preferred buffers are phosphate or tris-acetate at less than 0.1 M concentration. Because the overall method is based on detection of trace atoms, it is critically important that the buffers be free of contaminant ions, such as sulfate.

A second satisfactory method to create an aerosol of the column effluent is an ultrasonic device which produces sufficient mechanical shear and local heating to disperse the droplet. A similar type of shear may also be generated off the tip of a capillary injector into a venturri type aspirator, where the flow of supporting gas is developed by the pressure differential into the mass spectrometer. In these designs, it is often desirable to add small amounts of additional aqueous or organic solvents at the tip in order to aid in the flash evaporation of the effluent.

The combustion section is designed to completely burn the vaporized or surface evaporated chromatography effluent of DNA fragments together with the supporting electrolytes and optional solvent modifier to the oxides of carbon, hydrogen, nitrogen, phosphorous, and most importantly, those of the marker isotope. The sensitivity of detection of the marker isotope is greatly affected by the efficiency of combustion, since low molecular weight fragment ions which may result from incomplete burning can mask the presence of the primary detection species. The sequencing preparation must

be free of other ions in the mass/charge range of 64 to 68, since such ions would interfere with detection of the isotopic sulfur dioxide.

The combustion may be accomplished at moderate to approximately atmospheric pressure prior to injection into the mass spectrometer. For sulfur containing streams, essentially complete combustion to sulfur dioxide will be achieved when the sample is heated to temperatures in excess of approximately 900°C in an oxygen environment.

The most rugged design is to simply aspirate the column effluent into a hydrogen-oxygen flame, similar in design to standard gas chromatography flame ionization detectors. The important characteristics of the flame, the temperature and sample residence time, will be determined by the ratio of hydrogen to oxygen, the aggregate flow rate of gases and the local pressure. The characteristics of sulfur containing flames at 100-150 torr have been described by Zachariah and Smith, Combustion and Flame, 69, 125 (1987). A limitation to the sensitivity of this design is the volume of gas (water vapor) resulting from hydrogen-oxygen combustion which effectively dilutes the sulfur dioxide. Although the standard mass spectrometry techniques such as gas separators or semipermeable membranes may be used to remove water vapor, there is a trade-off between sample dilution and ultimate detectability which must be considered for each design.

A preferred method to very efficiently burn the nebulized column effluent is to inject it into a short heated tube in an oxygen environment. The tube may be constructed of nobel metals such as platinum, ceramic or quartz, depending on the method of heating. The external heating action may be provided by a cartridge electrical resistance heater or an external flame. The tube may be packed with a heat exchanger medium such as glass wool. Optionally, a catalytic surface may also be provided by such materials as supported platinum or copper oxide to enhance combustion efficiency.

A particularly effective method to burn the sample is in an inductively coupled oxygen plasma, where the tube forms the resonant cavity of a microwave generator. The inductively coupled plasma techniques have been reviewed by G. Meyer, Anal. Chem., 59, 1345A (1987).

Alternatively, the combustion may be affected within the low pressure environment of the mass spectrometer. A standard ionization technique in mass spectrometry is fast atom bombardment. An energetic beam of atoms,

usually xenon, is produced in a plasma torch and directed toward the sample. Ionization occurs by collision induced dissociation of this beam with the sample. If instead of pure xenon, oxygen is introduced into the fast atom beam, then both oxidation and ionization of the sample can occur. The limitation of this method, however, is the difficulty of achieving quantitative oxidation, and thus minimizing the background signal from incompletely oxidized low mass fragments.

There are several methods to effect ionization of sulfur dioxide. The objective is to obtain as high ion current as possible. In designs which operate at atmospheric pressure by flame, corona discharge needle or microwave induced plasma discharge, the ionization efficiency will be very high. In this type of design, a portion of the ionized gas is introduced into the low pressure region of the mass analyzer through a small sampling orifice or skimmer cone. The size of the orifice, and thus the percentage of total combustion sample which can be introduced, will depend on the pumping speed of the vacuum system. Generally, less than five percent of the sample will be transferred to the analyzer region. Designs of this type have been described by T. Covey, Anal. Chem., 58, 1451A (1986) and by G. Hieftje, Anal. Chem., 59, 1644 (1987).

Alternatively, the sample may be ionized in the lower pressure region near the analyzer. This may be achieved by such methods as electron impact using the beam emanating from a hot filament, by fast atom bombardment with inert gases such as xenon, or by chemical ionization with a variety of light gases. In these types of design, although the ionization efficiency is low relative to atmospheric methods, a greater percentage of those ions actually get to the analyzer section. The electron impact techniques have been described by A. Bandy, Anal. Chem., 59, 1196 (1987).

An RF-only quadrupole mass filter may be used to help separate low molecular weight combustion products (H_2O,N_2 and CO_2).

The analyzer and ion detector sections can be selected from several commercially available designs. The analyzer may be a quadrupole device where mass selection depends on the trajectory in a hyperbolic field, a field swept electromagnetic device with a single ion detector, or a permanent magnet device with an array of four ion detectors tuned to the isotopes of interest. The detector may be of single stage or ion multiplier design, although the latter type is preferred for highest sensitivity.

When the mass spectrometer is being used to detect isotopes of sulfur dioxide, as would be the case when dideoxy terminated thiophosphates are being utilized, the very high sensitivity is achieved when the polarity of the spectrometer is set to determine the positive ion spectrum. However, in glow discharge ionization, highest sensitivity is achieved when the spectrometer is operated in the negative ion mode.

When the mass spectrometer is being used to detect isotopes of chlorine or bromine, then maximum sensitivity will be achieved when the spectrometer polarity is reversed, and the negative ion spectrum is detected.

The labeled compounds of the present invention are prepared by conventional reactions employing commercially available isotopes. For example, the sulfur isotopes: ^{32}S, ^{33}S, ^{34}S and ^{36}S are commercially available as CS_2 or H_2S from the Department of Energy (Oak Ridge, Tenn. or Miamisburg, Ohio) at "isotopic enrichments" of 99.8%, 90.8%, 94.3% and 82.2%, respectively. Although the method of the present invention will provide satisfactory results with the "isotopically enriched" commercial products, it is preferred that the sulfur isotopes be at 99.5% enrichment to facilitate interpretation of the ion current v. time plots.

Enrichment techniques for sulfur are well known in the art. In particular, CS_2 can be further enriched by fractional distillation which takes advantage of the different boiling points of CS_2 conferred by the various sulfur isotopes. Alternatively, gaseous diffusion of SF_6 also can provide further enrichment of the sulfur isotope of interest. Thereafter, the isotopically enriched CS_2^*, H_2S^* or SF^*_6 are converted into the reagents described herein by techniques well known in the art.

The "halogen" isotopes ^{35}Cl, ^{37}Cl, ^{79}Br and ^{81}Br are also commercially available from the Department of Energy in either elemental form or as the corresponding halide salt at enrichments of 99%, 95%, 90% and 90%, respectively. These isotopes are used herein in their commercially available form.

As shown in Scheme F, labeled 2',3'-dideoxynucleotide-5-(O-1-thiophosphates) III are prepared by initially reacting a 2',3'-dideoxynucleoside I with isotopically enriched thiophosphoryl chloride (PS^*Cl_3) in triethyl phosphate, wherein by "S*" is meant a sulfur isotope that is a member of the group consisting of ^{32}S, ^{33}S, ^{34}S and ^{36}S in isotopically enriched form. From the above reaction, the correspondingly

2',3'-dideoxynucleotide-5'-(O-1-thiotriphosphate) III is prepared from II by dissolving the bistriethylamine (TEA) salt of II in dioxane and reacting it with diphenyl phosphochloriodate to form the diphenyl phosphate ester of II. This phosphate ester was further reacted with the tetrasodium salt of pyrophosphate in pyridine to form III. Purification of III is accomplished by chromatography on diethylaminoethyl (DEAE) cellulose.

Scheme G shows the general method for preparing 3'-halo-2',3'-dideoxynucleosides V wherein the halogen ("X*") is a member of the group consisting of ^{35}Cl, ^{37}Cl, ^{79}Br, and ^{81}Br in isotopically enriched form. In particular, a solution of 1-(5-O-triphenylmethyl-2-deoxy-β-D-threopentofuranosyl)nucleoside wherein by nucleoside is meant A, T, G, or C, in a basic solvent, such as pyridine, was reacted with methanesulfonyl chloride. The resulting mesylate was treated with an isotopically enriched salt, such as LiX*, in the presence of heat and then acidified to produce a 3'-halo-2',3'-dideoxynucleoside V. Compounds of Formula V can be converted to the corresponding labeled nucleotide monophosphate VI by reaction with cyanoethyl phosphate and dicyclohexylcarbodiimide (DCC) followed by LiOH deblocking.

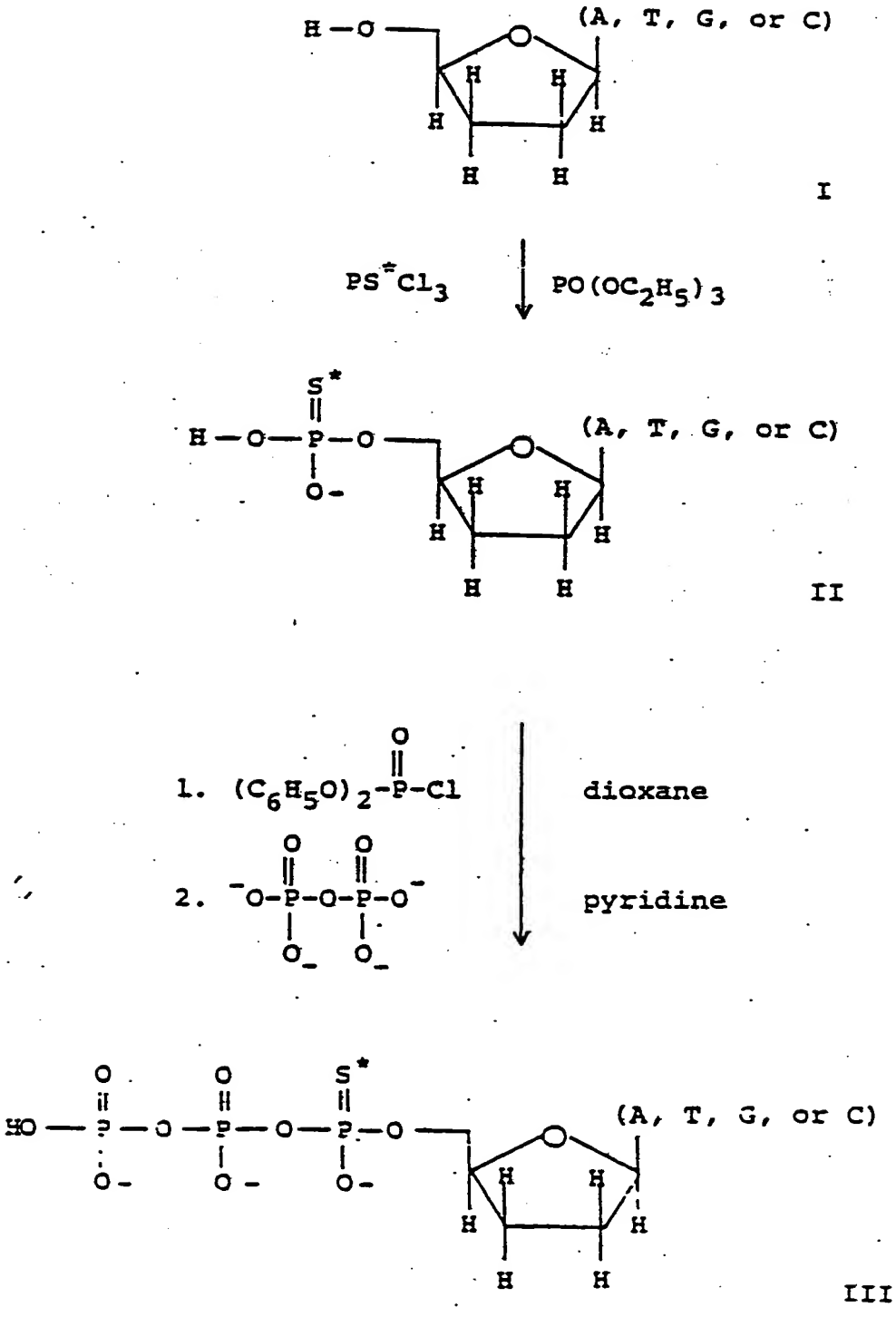
The corresponding triphosphate of V is prepared from the monophosphate as described from the monophosphate as described for the conversion of II to III above.

Scheme H presents a method for halogenating a purine or pyrimidine base of a 2',3'-dideoxynucleoside VII using isotopically enriched elemental bromine (^{79}Br or ^{81}Br) or chlorine (^{35}Cl or ^{37}Cl), i.e. X_2*. The 2',3'-dideoxynucleoside VII is dissolved in a polar solvent, such as dry DMF, in the presence of a base, such as pyridine. To the reaction mixture is added a molar equivalent of the elemental halogen (X_2*) and the reaction mixture is allowed to stir for 12 hours. Evaporation of the solvent produces the labeled 2',3'-dideoxynucleoside VIII wherein the isotopic halogen label is on the purine or pyrimidine base of the dideoxy nucleoside. Purification is accomplished by conventional chromatographic techniques.

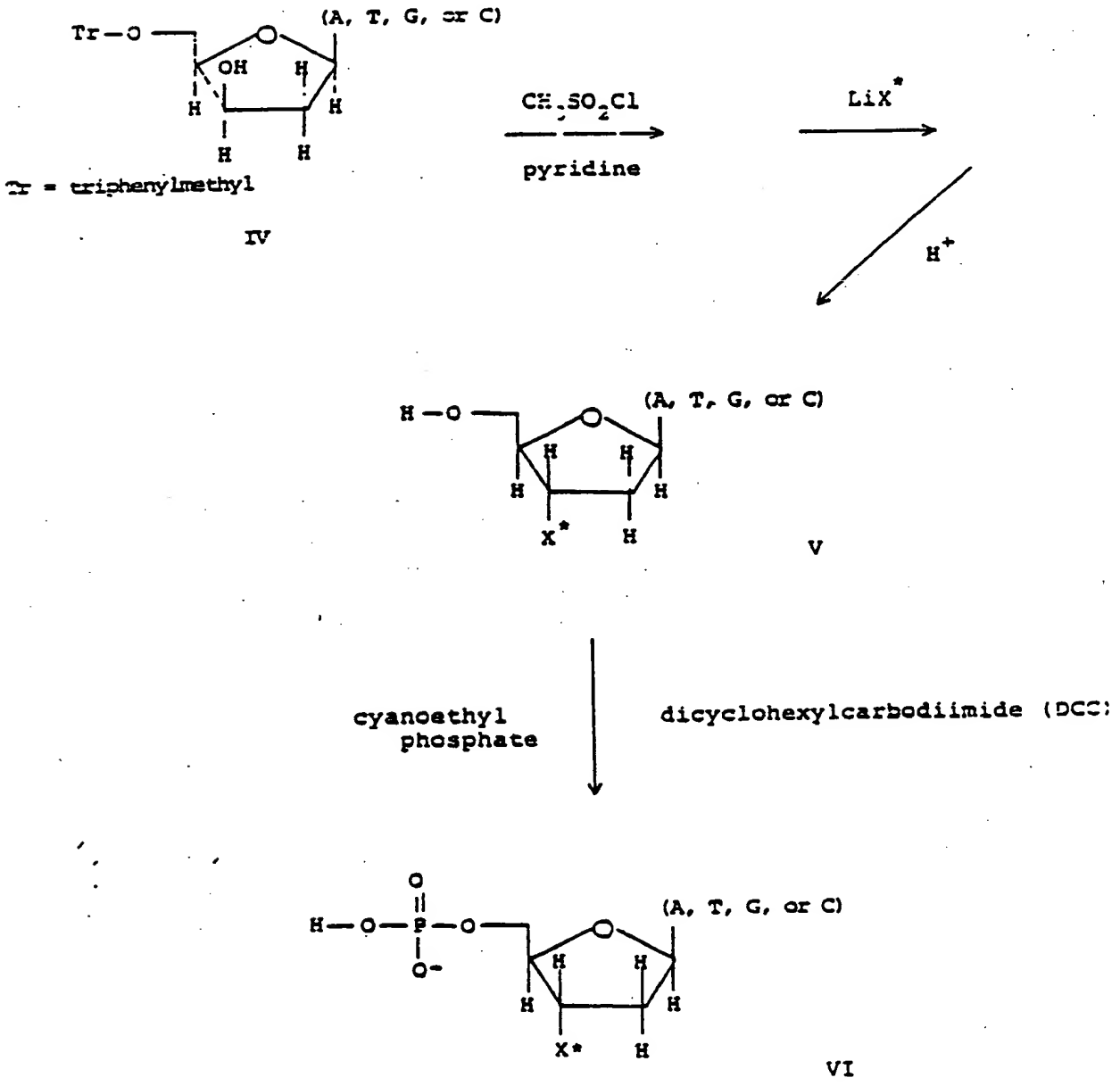
The labeled 2',3'-dideoxynucleoside VIII is converted to the corresponding monophosphate and triphosphate as discussed above for VI and III, respectively.

The examples described herein are intended to illustrate the present invention and not limit it in spirit or scope.

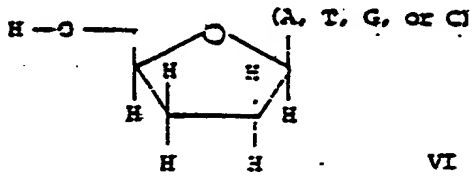
SCHEME F


(A, T, G, or C)
I
PS*Cl3
PO(OC2H5)3
(A, T, G, or C)
II
1. (C6H5O)2-P-Cl
dioxane
2.
pyridine
(A, T, G, or C)
III

SCHEME G

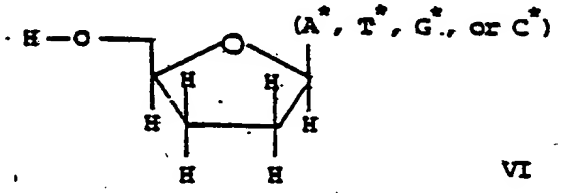

Tr—O
(A, T, G, or C)
OH
H
H
H
H
H
Tr = triphenylmethyl
IV
CH_3SO_2Cl
pyridine
LiX^*
H^+
H—O
(A, T, G, or C)
H
H
H
H
X^*
H
V
cyanoethyl phosphate
dicyclohexylcarbodiimide (DCC)
O
H—O—P—O
O-
(A, T, G, or C)
H
H
H
H
X^*
H
VI

SCHEME H



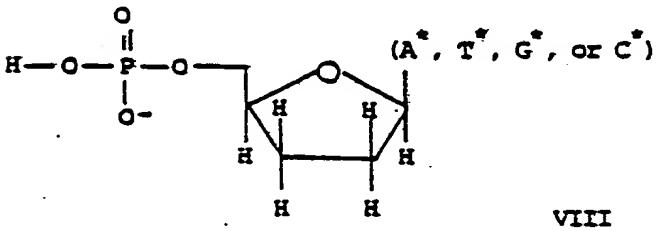
VI

X_2^* | pyridine



VI

cyanoethyl phosphate | dicyclohexylcarbodiimide (DCC)



VIII

EXAMPLE 1

Preparation of [^{32}S]2',3'-Dideoxyadenosine-5'-Phosphorothioate

2'3'-Dideoxyadenosine (47 mg, 0.2 mmol) was suspended in triethyl phosphate (0.5 ml) and heated briefly to 100°C. The solution was cooled to 4°C and treated with [^{32}S] thiophosphoryl chloride (37 mg, 0.22 mmol). The mixture was agitated for 12 hr. at 4°C, and then treated with 2 ml 10% barium acetate and agitated at 20°C for 1 hour. The suspension was treated with 0.5 ml triethyl amine and then with 5 ml 95% ethanol. The suspension was agitated for 30 min. and then filtered. The precipitate was washed with 50% aqueous ethanol and then water. The filtrate was evaporated to dryness, and the solid taken up in water and chromatographed on a column of diethylaminoethyl (DEAE) cellulose which had been equilibrated with NH_4HCO_3. The column was eluted with 0.1M NH_4HCO_3 and the fractions adsorbing at 260 nm were pooled and evaporated. The solid was evaporated twice with 80% ethanol, twice with 80% ethanol containing 2% triethyl amine (TEA), and finally with anhydrous ethanol. There was obtained 44 mg of the bis-triethylamine salt of the title product. A solution of the triethylamine salt in 1 ml methanol was treated with 1 ml of a solution of 6M NaI in acetone. The precipitate was washed with acetone and dried to give 32 mg of disodium salt of the title product as a white solid.

EXAMPLE 2

Preparation of [^{32}S]2',3'-Dideoxyadenosine-5'-(O-1-Thiotriphosphate)

A solution of the bis-triethylamine (TEA) salt of the title product of Example 1 (26 mg, 0.05 mmol) was dissolved in 1 ml dry dioxane and treated with diphenyl phosphochloridate (0.015 ml, 0.075 mmol). The mixture was agitated for 3 hr. at 25°C. A solution of dry pyrophosphate in pyridine was prepared by dissolving the tetrasodium salt (220 mg, 0.5 mmol) in 3 ml pyridine and evaporating twice, and the taking up in 0.5 ml pyridine. This solution was added to the above solution of the crude active ester, and stirred for 2 hr. The crude product was precipitated by addition of ether (10 ml). The precipitate was dissolved in water and chromatographed in DEAE-cellulose eluted with 0.1M triethylammonium bicarbonate. The pooled fractions contained 150 A_{260} units (20%) of the title product. The solution was lyophilized and the residue stored at -70°C.

EXAMPLE 3

Preparation of [^{79}Br]3'-Bromo-2',3'-Deoxythymidine

A solution of 1-(5-O-triphenylmethyl-2'-deoxy-β-D-threopentofuranosyl)thymine (50 mg, 0.1 mmol) in pyridine (1ml) was treated with methanesulfonyl chloride (0.014 ml, 0.12 mmol) and the reaction agitated at 10°C for 6 hr. The mixture was then evaporated, diluted with $CHCl_3$ (5 ml) and washed 2x with water. The organic layer was evaporated, and the crude mesylate dissolved in dry diglyme (1 ml) and treated with [Li^{79}Br] (17 mg, 0.2 mmol). The solution was heated for 4 hr. at 100°C, and then diluted with 80% acetic acid (1 ml) and heated 15 min. longer. The reaction was cooled and diluted with water (2 ml), and extracted 3x with chloroform (2 ml). The organic extracts were evaporated and the residue chromatographed on silica gel eluted with 95:5 chloroform methanol to give 18 mg of the title product as an off-white solid.

This material could be converted to the triphosphate via the monophosphate as in Example 2. The monophosphate was prepared by reaction with cyanoethyl phosphate and dicyclohexylcarbodiimide (DCC), followed by LiOH deblocking.

EXAMPLE 4

Preparation of [^{79}Br]2',3'-Dideoxy-5-Bromocytidine

To a solution of 2',3'-dideoxycytidine (60 mg, 0.3 mmol) in dry DMF (1 ml) was added 0.1 ml pyridine and then [^{79}Br] bromine (42 mg, 0.3 mmol), and the mixture agitated for 12 hr. The solvent was evaporated, and the residue chromatographed on silica gel (ethyl acetate:methanol:triethylamine 90:10:1) to give 46 mg of the title product as a white solid.

This material could be converted to the triphosphate via the monophosphate as in Example 2. The monophosphate was prepared by reaction with cyanoethyl phosphate and dicyclohexylcarbodiimide followed by LiOH deblocking.

What is Claimed is:

1. A method for determining the base sequence of natural or artificially made DNA or RNA or fragments thereof, a step in said method comprising:

determining the identity of the bases in said sequence by identifying the unique stable foreign isotope associated with each variety of base.

2. In a method to determine a DNA sequence by the dideoxynucleotide chain termination method, the improvement comprising identifying the terminal nucleotide of the chain terminated DNA sequences by measuring by mass spectrometry an isotope specifically incorporated in the chain terminated DNA sequence, whereby the mass of the incorporated isotope or its oxidation or combustion product corresponds with the identity of specific terminal nucleotide in each said terminated DNA sequence.

3. A method according to Claim 2 wherein the isotope is incorporated in a dideoxynucleotide.

4. A method according to Claim 2 wherein the isotope is incorporated into a deoxynucleotide.

5. A method according to Claim 2 wherein the isotope is incorporated into a DNA primer.

6. An improved method for sequencing DNA wherein deoxynucleotide triphosphates, consisting of dTTP, dATP, dCTP, and dGTP, and 2,3-dideoxynucleotide triphosphates, consisting of ddATP, ddCTP, ddGTP, and ddTTP, are reacted with a DNA sequence in the presence of a DNA primer which binds to the DNA sequence and a DNA polymerase; whereby complementary DNA sequences are formed by the addition of complementary nucleotides to the primer and are terminated by dideoxynucleotide triphosphate incorporation providing a terminated complementary DNA sequence corresponding with each of the bases in the DNA sequence to be determined and separating the resulting terminated DNA sequences according to size; the improvement comprising;

labeling ddATP, ddCTP, ddGTP, and ddTTP respectively with stable isotopes of different mass, and

detecting the isotope of different mass in each of the terminated DNA sequences by mass spectroscopy so as to identify the terminal

nucleotide in each said terminated complementary DNA sequence, thereby determining the sequence of the DNA;

whereby when the identity of the terminal nucleotide is plotted versus the size of the fragment, the DNA sequence of the complementary DNA is determined.

7. A method according to Claim 6 wherein ^{32}S, ^{33}S, ^{34}S and ^{36}S isotopes are used.

8. A DNA sequence analyzer comprising

(a) a means for separating chain terminated DNA sequences according to size;

(b) a mass spectrometer for analyzing isotopes of different mass bound to the chain terminated DNA sequence in a relationship that associates the mass of an isotope with a terminal nucleotide of the DNA sequence; and

(c) a means for transporting the chain terminated DNA sequences from the means for separating the chain terminated DNA sequences to a combustion chamber with converts the elements of chain terminated DNA sequence including the isotopes to oxides for analysis in the mass spectrometer.

9. A DNA sequence analyzer according to Claim 8 wherein the mass spectrometer contains a means for converting ^{32}S, ^{33}S, ^{34}S and ^{36}S to $^{32}SO_2$, $^{33}SO_2$, $^{34}SO_2$ and $^{36}SO_2$ and sequentially detecting the SO_2 resulting from the chain terminated DNA sequences.

10. A reagent for DNA sequencing comprising a mixture of dideoxynucleotide triphosphates (ddCTP, ddGTP, ddTTP, and ddATP) wherein each dideoxynucleotide triphosphate is labeled with a different isotope which is measurable by mass spectrometry.

11. A mixture according to Claim 10 wherein the isotope is sulfur isotope selected from ^{32}S, ^{33}S, ^{34}S or ^{36}S.

12. A mixture of dideoxynucleotide chain terminated DNA sequences wherein each chain terminated DNA sequence contains an isotope measurable by mass spectrometry which is specifically related to a specific chain terminating dideoxynucleotide.

13. A mixture according to Claim 12 wherein the isotope is ^{32}S, ^{33}S, ^{34}S or ^{36}S.

14. A mixture according to Claim 12 wherein at least one of the isotopes is a halogen.

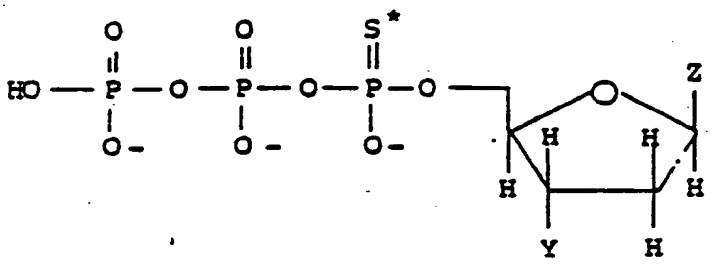
15. A mixture according to Claim 12 which is separated according to size.

16. A mixture according to Claim 12 wherein the isotope is in the DNA primer portion of the terminated DNA sequence.

17. A mixture according to Claim 12 wherein the isotope is in the dideoxynucleotide portion of the chain terminated DNA sequence.

18. A mixture according to Claim 12 wherein the isotope in the chain terminated DNA sequence originates with the deoxynucleotide triphosphate.

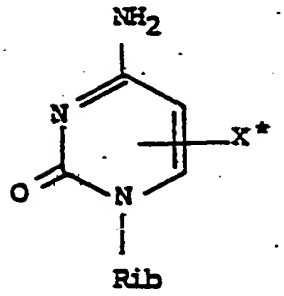
19. A compound of the formula:



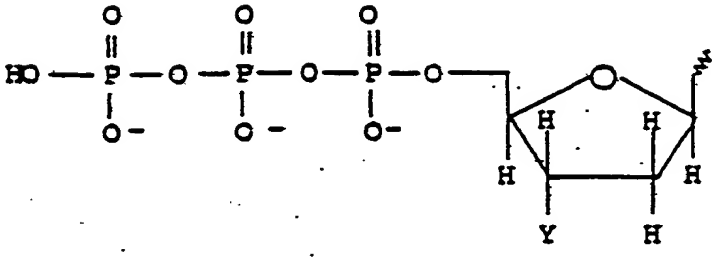
wherein S* is selected from ^{32}S, ^{33}S, ^{34}S or ^{36}S; wherein Z is adenine, cytosine, guanine or thymidine; and wherein Y is hydrogen or hydroxyl.

20. A mixture of the compounds of Claim 19 wherein one sulfur isotope is associated with one of A, T, G, or C and Y is hydrogen.

21. A reagent for DNA sequencing according to the formula:

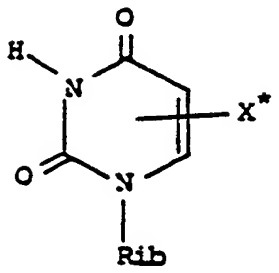


where X is halogen or lower alkylthio having 1-6 carbon atoms; wherein Rib is

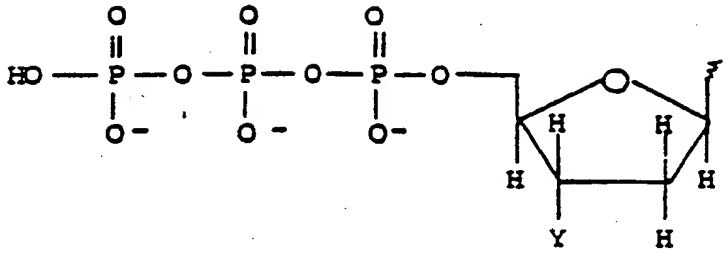


wherein Y is hydrogen or hydroxyl; and wherein the sulfur of said alkylthio or said halogen is isotopically enriched and detectable by mass spectroscopy.

22. A reagent for DNA sequencing according to the formula:

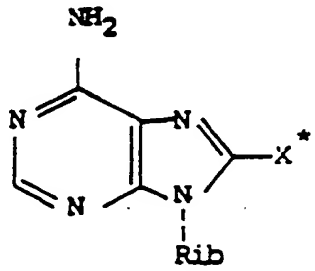


where X is halogen or lower alkylthio having 1-6 carbon atoms; wherein Rib is

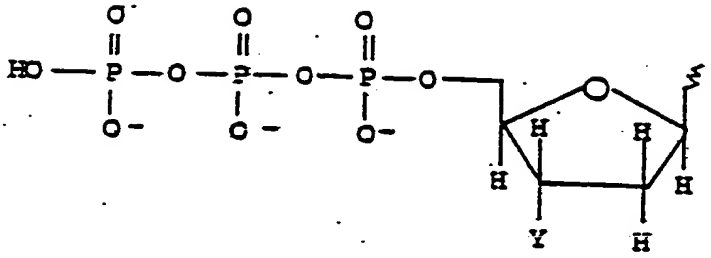


wherein Y is hydrogen or hydroxyl; and wherein the sulfur of said alkylthio or said halogen is isotopically enriched and detectable by mass spectroscopy.

23. A reagent for DNA sequencing according to the formula:

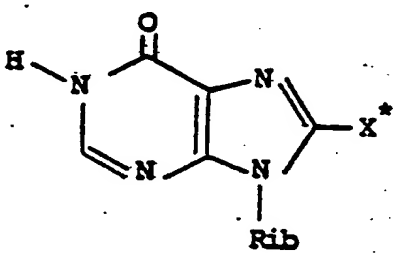


where X is halogen or lower alkylthio having 1-6 carbon atoms; wherein Rib is

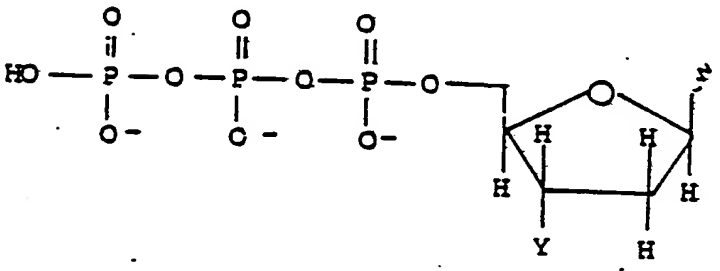


wherein Y is hydrogen or hydroxyl; and wherein the sulfur of said alkylthio or said halogen is isotopically enriched and detectable by mass spectroscopy.

24. A reagent for DNA sequencing according to the formula:



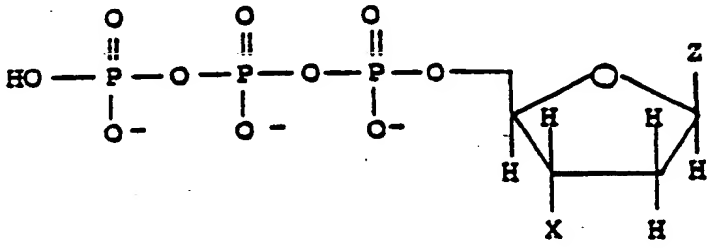
wherein X is halogen or lower alkylthio having 1-6 carbon atoms; wherein Rib is



wherein Y is hydrogen or hydroxyl; and wherein the sulfur from said alkylthio or said halogen is isotopically enriched and detectable by mass spectroscopy.

25. A reagent according to Claims 21, 22, 23 or 24 wherein sulfur is selected from ^{32}S, ^{33}S, ^{34}S and ^{36}S, and wherein halogen is selected from ^{35}Cl, ^{37}CL, ^{79}Br, and ^{81}Br.

26. A reagent for DNA sequencing according to the formula:



wherein Z is adenine, guanine, cytosine, or thymine;
wherein X is ^{35}Cl, ^{37}Cl, ^{79}Br or ^{81}Br.

Fig. 1

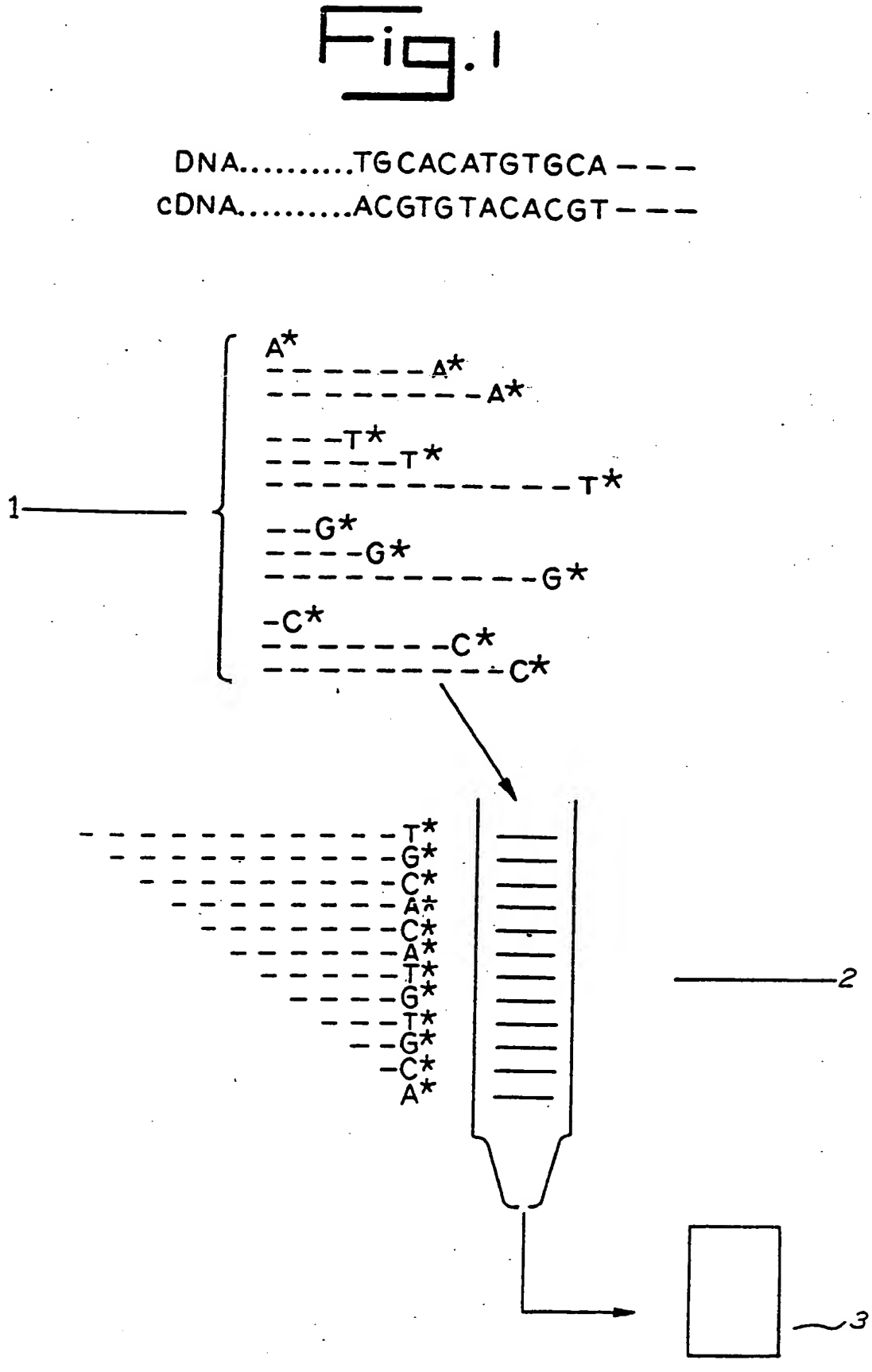


Fig. 2A

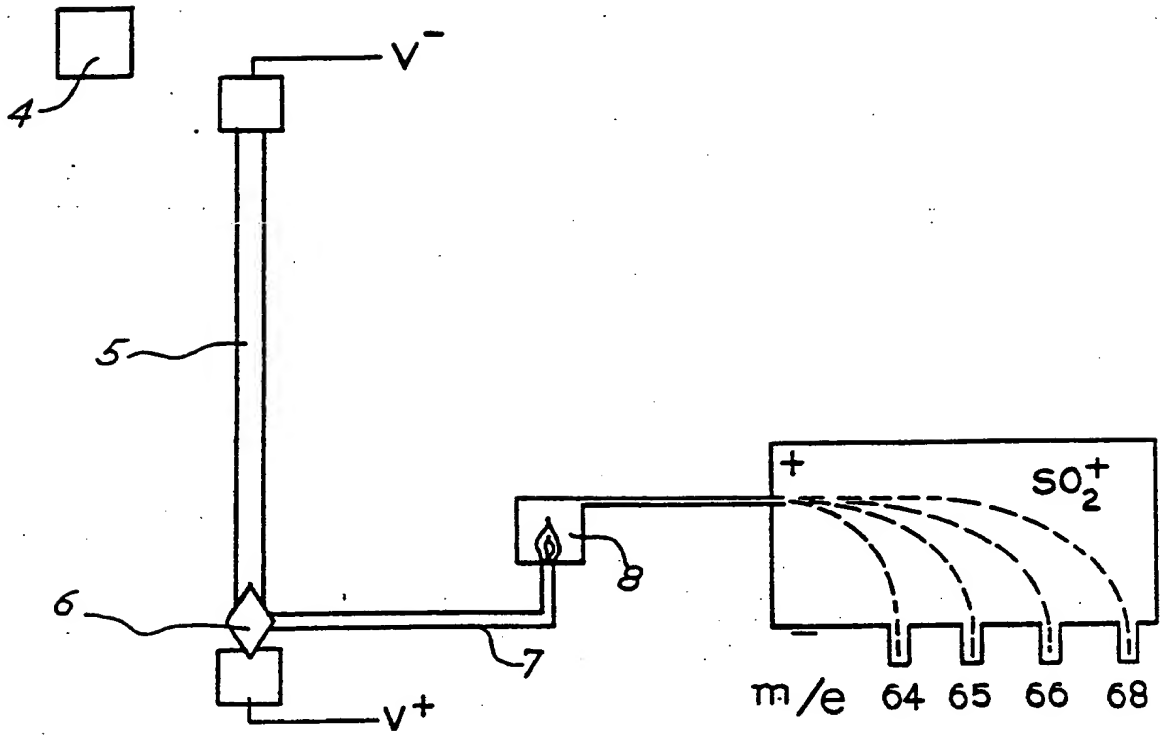
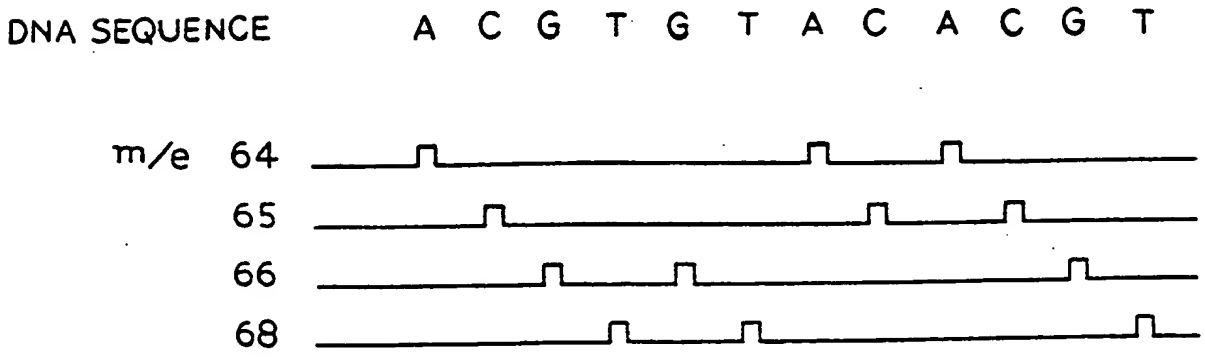


Fig. 2B



SO_2 ION CURRENT V. TIME

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 89/02602

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC [4]: C 12 Q 1/68, C 07 H 19/00

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System [8]	Classification Symbols
IPC4	C 12 Q

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [9]

III. DOCUMENTS CONSIDERED TO BE RELEVANT [10]

Category [11]	Citation of Document, [12] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
A	EP, A, 0223618 (NEW YORK MEDICAL COLLEGE) 27 May 1987 see column 1, line 1 - column 5, line 34; column 11, lines 30-36 --	1-7
A	DE, A, 3312929 (GESELLSCHAFT FÜR BIO-TECHNOLOGISCHE FORSCHUNG mbH) 8 December 1983 see page 1, line 1 - page 4, line 6; claims --	1
A	Biological Abstracts, volume 70, no. 12, (Philadelphia, PA, US), P.E. Agris et al.: "High performance liquid chromatography and mass spectrometry of transfer RNA bases for isotopic abundance", see page 7947, abstract 76072, & J. Chromatogr. 194(2): 205-216, 1980 --	1
A	Accounts of Chemical Research, volume 12, 1979, American Chemical Society, ./.	19

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
4th October 1989	08 NOV 1989
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	T.K. WILLIS

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
	F. Eckstein: "Phosphorothioate analogues of nucleotides", pages 204-210; see pages 204-205 cited in the application	
	--	
A	Biochemistry, volume 16, no. 6, 1977, J.L Wiebers et al.: "Sequence analysis of oligodeoxyribonucleotides by mass spectrometry. 1. Dinucleoside monophosphates", pages 1044-1050 see abstract	1

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. US 8902602 SA 29529

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 31/10/89
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A- 0223618	27-05-87	JP-A- 62085863	20-04-87
DE-A- 3312929	08-12-83	EP-A,B 0103677	28-03-84
		JP-A- 59026064	10-02-84

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US88/03194

I. CLASSIFICATI N OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC C12Q 1/68
U.S. CL.: 435/6

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
U.S.	435/6,41 935/77,78,88 436/501,94, 175, 800

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
X	US, X, 3,730,844 GILHAM ET AL, 01 May 1973,	1-21
Y	US, Y, 4,521,509, BENKOVIC ET AL, 04 June 1985,	1-21
Y	Nucleic Acids Research, Vol. 14, No. 6, 1986, Harvard University, Baumlein et al, 'The Legumin Gene Family: Structure of a B Type Gene of Vicia Faba and A Possible Legumin Gene Specific Regulatory Element', pp. 2707-2720.	1-21
Y	DNA, Vol. 5, No. 3, 1986, Medical College of Wisconsin, Milwaukee, WI, SHIMKUS ET AL, 'Synthesis and Characterization of Biotin-Labeled Nucleotide Analogs', pp. 247-255.	1-21

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
12 January 1989	07 MAR 1989
International Searching Authority	**Signature of Authorized Officer**
ISA/US	AMELIA BURGRESS YARBROUGH

Form PCT/ISA/210 (second sheet) (Rev.11-87)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
Y	Nature, Vol. 321, 12 June 1986, Foster City, Calif, Smith et al, "Fluorescence Detection in Automated DNA sequence Analysis,' pp. 674-679.	1-21
Y	Proc. Natl. Acad. Sci, Vol. 78, No. 11, New Haven, Conn, Langer et al "Enzymatic Synthesis of Biotin-Labeled Polynucleotides: Novel Nucleic Acid Affinity Probes', pp. 6633-6637, November 1981.	1-21
Y,P	EU, Y, 0251,575 Houghton et al, 07 January 1988, pp. 10-11.	1-21

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 : C12Q 1/68	A1	(11) International Publication Number: WO 89/ 03432
		(43) International Publication Date: 20 April 1989 (20.04.89)

(21) International Application Number: PCT/US88/03194

(22) International Filing Date: 16 September 1988 (16.09.88)

(31) Priority Application Number: 105,375

(32) Priority Date: 7 October 1987 (07.10.87)

(33) Priority Country: US

(71) Applicant: UNITED STATES DEPARTMENT OF ENERGY [US/US]; 1000 Independence Avenue, S.W., Washington, DC 20585 (US).

(72) Inventors: JETT, James, Hubert ; 545 Navajo Road, Los Alamos, NM 87544 (US). KELLER, Richard, Alan ; 4 La Rosa Court, Los Alamos, NM 87544 (US). MARTIN, John, Calvin ; 4987 Trinity Drive, Los Alamos, NM 87544 (US). MOYZIS, Robert, Keith ; 865 Los Pueblos, Los Alamos, NM 87544 (US). SHERA, Edgar, Brooks ; 4 Timber Ridge Road, Los Alamos, NM 87544 (US). STEWART, Carleton, Colburn ; 30 Carlton Drive, Hamburg, NY 14075 (US). RATLIFF, Robert, Lafayette ; 252 La Cueva, Los Alamos, NM 87544 (US).

(74) Agents: CONSTANT, Richard, E. et al.; United States Department of Energy, GC-42 (FORSTL) MS 6F-067, 1000 Independence Avenue, S.W., Washington, DC 20585 (US).

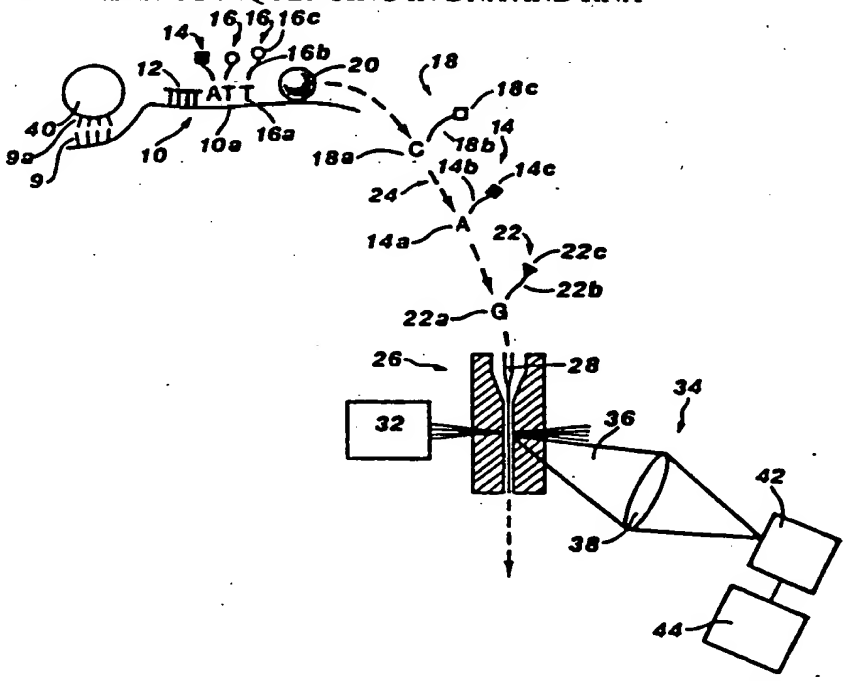
(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), NO, SE (European patent).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD FOR RAPID BASE SEQUENCING IN DNA AND RNA



(57) Abstract

A method is provided for the rapid base sequencing of DNA or RNA fragments wherein a single fragment of DNA or RNA (40) is pr vided with identifiable bases (A, C, G) and suspended in a moving flow stream. An exonuclease (20) sequentially cleaves individual bases from the end of the suspended fragment. The moving flow stream (24) maintains the cleaved bases in an orderly train for subsequent detecti n and identification. In a particular embodiment, individual bases forming the DNA or RNA fragments are individually tagged with a characteristic fluorescent dye. The train f bases is then excited to fluorescence with an output spectrum characteristic f the individual bases (26). Accordingly, the base sequence of the original DNA or RNA fragment can be reconstructed.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BG	Bulgaria
BJ	Benin
BR	Brazil
CF	Central African Republic
CG	Congo
CH	Switzerland
CM	Cameroon
DE	Germany, Federal Republic of
DK	Denmark
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
HU	Hungary
IT	Italy
JP	Japan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
MC	Monaco
MG	Madagascar
ML	Mali
MR	Mauritania
MW	Malawi
NL	Netherlands
NO	Norway
RO	Romania
SD	Sudan
SE	Sweden
SN	Senegal
SU	Soviet Union
TD	Chad
TG	Togo
US	United States of America

METHOD FOR RAPID BASE SEQUENCING IN DNA AND RNA

BACKGROUND OF THE INVENTION

This invention is generally related to DNA and RNA sequencing and, more particularly, to DNA and RNA sequencing by detecting individual nucleotides. This invention is the result of a contract with the Department of Energy (Contract No. W-7405-ENG-36).

A world-wide effort is now in progress to analyze the base sequence in the human genome. The magnitude of this task is apparent, with 3×10^9 bases in the human genome, and available base sequencing rates are about 200-500 bases per 10-24 hour period. Considerable interest also exists in nucleic acid sequencing from non-human sources. Existing procedures are labor intensive and cost approximately $1 per base.

By way of example, Sanger et al., "DNA Sequencing with Chain-Terminating Inhibitors," Proceedings of the National Academy of Science, USA 74, 5463-7 (1977) provide for sequencing 15-200 nucleotides from a priming site. Radioactive phosphorus is used in the primer extension to provide a marker. Enzymatic resynthesis coupled with chain terminating precursors are used to produce DNA fragments which terminate randomly at one of the four DNA bases: adenine (A), cytosine (C), guanine (G), or thymine (T). The four sets of reaction products are s parated

electrophorectically in adjacent lanes of a polyacrylamide gel. The migration of the DNA fragments is visualized by the action of the radioactivity on a photographic film. Careful interpretation of the resulting band patterns is required for sequence analysis. This process typically takes 1-3 days. Further, there are problems with band pile-ups in the gel, requiring further confirmatory sequencing.

In a related technique, A.M. Maxam and W. Gilbert, "A New Method for Sequencing DNA," Proceedings of the National Academy of Science, USA 74, 560-564 (1977), teach a chemical method to break the DNA into four sets of random length fragments, each with a defined termination. Analysis of the fragments proceeds by electrophoresis as described above. The results obtained using this method are essentially the same as the "Sanger Method."

In another example, Smith et al., "Fluorescent Detection in Automated DNA Sequence Analysis," Nature 321, 674-679 (June 1986), teach a method for partial automation of DNA sequence analysis. Four fluorescent dyes are provided to individually label DNA primers. The Sanger method is used to produce four sets of DNA fragments which terminate at one of the four DNA bases with each set characterized by one of the four dyes. The four sets of reaction products, each containing many identical DNA fragments, are mixed together and placed on a polyacrylamide gel column. Laser excitation is then used to identify and characterize the migration bands of the labeled DNA fragments on the column where the observed spectral properties of the fluorescence are used to identify the terminal base on each fragment. Sequencing fragments of up to 400 bases has been reported. Data

reliability can be a problem since it is difficult to uniquely discern the spectral identity of the fluorescent peaks.

These and other problems in the prior art are addressed by the present invention and an improved process is provided for rapid sequencing of DNA bases. As herein described, the present invention provides for the sequential detection of individual nucleotides cleaved from a single DNA or RNA fragment.

Accordingly, it is an object of the present invention to provide an automated base sequence analysis for DNA and RNA.

Another object of the present invention is to process long strands of DNA or RNA, i.e., having thousands of bases.

One other object is to rapidly sequence and identify individual bases.

Additional objects, advantages and novel features of the invention will be set forth in part in the description which follows, and in part will become apparent to those skilled in the art upon examination of the following or may be learned by practice of the invention. The objects and advantages of the invention may be realized and attained by means of the instrumentalities and combinations particularly pointed out in the appended claims.

SUMMARY OF THE INVENTION

To achieve the foregoing and other objects, and in accordance with the purposes of the present invention, as embodied and broadly described herein, a method for DNA and RNA base sequencing is provided. A single fragment from a strand of DNA or RNA is suspended in a moving sample stream. Using an exonuclease, the end base on the

DNA or RNA fragment is repetitively cleaved from the fragment to form a train of the bases in the sample stream. The bases are thereafter detected in sequential passage through a detector to reconstruct the base sequence of the DNA or RNA fragment.

In another characterization of the present invention, strands of DNA or RNA are formed from the constituent bases, which have identifiable characteristics. The bases are sequentially cleaved from the end of a single fragment of the strands to form a train of the identifiable bases. The single, cleaved bases in the train are then sequentially identified to reconstruct the base sequence of the DNA or RNA strand.

In one particular characterization of the invention, each of the nucleotides effective for DNA and RNA resynthesis is modified to possess an identifiable characteristic. A strand of DNA is synthesized from the modified nucleotides, where the synthesized strand is complementary to a DNA or RNA strand having a base sequence to be determined. A single fragment of the complementary DNA or RNA is selected and suspended in a flowing sample stream. Individual identifiable nucleotides are sequentially cleaved from the free end of the suspended DNA strand. The single bases are then sequentially identified. The base sequence of the parent DNA or RNA strand can then be determined from the complementary DNA strand base sequence.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and form a part of the specification, illustrate an embodiment of the present invention and, together with the description, serve to explain the principles of the invention. In the drawings:

FIGURE 1 is a graphic illustration of a DNA sequencing process according to the present invention.

FIGURE 2 is a graphical representation of an output signal according to the present invention.

DETAILED DESCRIPTION OF THE INVENTION

According to the present invention, a method is provided for sequencing the bases in large DNA or RNA fragments by isolating single DNA or RNA fragments in a moving stream and then individually cleaving single bases into the flow stream, forming a sequence of the bases through a detection device. In one embodiment, the single bases in the flowing sample streams are interrogated by laser-induced fluorescence to determine the presence and identity of each base.

It will be understood that DNA and RNA strands are each formed from nucleotides comprising one of four organic bases: adenine, cytosine, guanine, and thymine (DNA) or uracil (RNA). The DNA and RNA nucleotides are similar, but not identical; however, the nucleotides and strands of nucleotides can be functionally manipulated in a substantially identical manner. Also, the complement of an RNA fragment is conventionally formed as a DNA strand with thymine in place of uracil. The following description is referenced to DNA sequencing, but any reference to DNA includes reference to both DNA and RNA and without any limitation to DNA.

In a particular embodiment of the present invention, the initial step is an enzymatic synthesis of a strand of DNA, complementary to a fragment to be sequenced, with each base containing a fluorescent tag characteristic of the base. Sequencing the complementary strand is equivalent to sequencing the original fragment. The synthesized strand is then suspended in a flowing sample

stream containing an exonuclease to cleave bases sequentially from the free end of the suspended DNA or RNA. The cleaved, fluorescently labeled bases then pass through a focused laser beam and are individually detected and identified by laser-induced fluorescence.

The maximum rate that bases may be sequenced is determined by the kinetics of the exonuclease reaction with DNA or RNA and the rate of detection. A projected rate of 1000 bases/sec would result in sequencing 8×10^7 bases/day. This is in contrast to standard techniques which take 10-24 hours to sequence 200-500 bases.

Referring now to Figure 1, one effective sequencing method comprises the following steps: (1) prepare a selected strand of DNA 10, in which individual bases are provided with an identifiable characteristic, e.g., labeled with color-coded fluorescent tags to enable each of the four bases to be identified, (2) select and suspend 40 a single fragment of DNA with identifiable bases in a flowing sample stream, (3) sequentially cleave 20 the identifiable bases from the free end of the suspended DNA fragment, and (4) identify the individual bases in sequence, e.g., detect 34 the single, fluorescently labeled bases as they flow through a focused laser system. Exemplary embodiments of the individual process steps are hereinafter discussed.

Selection of DNA Fragment to be Sequenced

In accordance with the present process, a single DNA fragment 10a is selected and prepared for labeling and analysis. In an exemplary selection process from a heterogeneous mixture of DNA fragments, avidin is bound to microspheres and a biotinylated probe, complementary to some sequence within the desired DNA fragment 10a, is

bound to the avidin on the microspheres. The avidin-biotinylated probe complex is then mixed with the heterogeneous mixture of DNA fragments to hybridize with the desired fragments 10a. The beads are separated from the unbound fragments and washed to provide the desired homogeneous DNA fragments 10a.

The selected fragments are further processed by removing the first microsphere and ligating a tail of known sequence 9 to the primer 12 attached to the 3' end of the fragment 10a. Microspheres 40 are prepared with phycoerythrin-avidin and sorted to contain a single molecule of phycoerythrin-avidin. A single complementary probe 9a to the known sequence 9 is biotinylated and bound to the sorted microspheres 40. The bead-probe complex is then hybridized to the selected fragment 10a. Thus, a single fragment of DNA 10a will be bound to each microsphere.

In another embodiment, a homogeneous source of DNA fragments is provided, e.g. from a gene library. A selection step is not then required and the homogeneous DNA fragments can be hybridized with the microspheres 40 containing a single molecule of phycoerythrin-avidin, with the appropriate complementary probe attached as above.

In either case, a single microsphere 40 can now be manipulated using, for example, a microinjection pipette to transfer a single fragment strand for labeling and analysis as discussed below.

Fluorescence Labeling of Bases

The bases forming the single fragment to be analyzed are provided with identifiable characteristics. The identifiable characteristic may attach directly to each nucleotide of DNA strand 10a. Alternatively, bases may first be modified to obtain individual identifiable

characteristics and resynthesized to selected strand 10a to form a complementary DNA strand. In either event, DNA fragment 10 is provided for analysis with identifiable bases.

In one embodiment, a fluorescent characteristic is provided. The bases found in DNA do have intrinsic fluorescence quantum yields $<10^{-3}$ at room temperature. In order to detect these bases by a fluorescence technique, however, it is desirable to modify them to form species with large fluorescence quantum yields and distinguishable spectral properties, i.e., to label the bases.

It is known how to synthesize a complementary strand of DNA with labeled bases using an enzymatic procedure. See, e.g., P. R. Langer et al., "Enzymatic Synthesis of Biotin-Labeled Polynucleotides: Novel Nucleic Acid Affinity Probes," Proc. Natl. Aca. Sci. USA 78, 6633 (1981); M. L. Shimkus et al., "Synthesis and Characterization of Biotin-Labeled Nucleotide Analogs," DNA 5, 247 (1986); all incorporated herein by reference. Referring to Figure 1, a primer 12 is attached to the 3' end of a DNA fragment 10a and an enzyme, e.g., DNA polymerase-Klenow fragment, is used to synthesize the complement to DNA fragment 10a starting from the end of primer 12. Modified deoxynucleotides 14, 16, 18, 22 are used in the synthesis (typically modified dATP 14a, dTTP (or dUTP) 16a, dCTP 18a, and dGTP 22a).

Each of the modified nucleotides is formed with a long carbon chain linker arm 14b, 16b, 18b, and 22b, respectively, terminating in a characteristic fluorescent dye 14c, 16c, 18c, and 22c. The modified nucleotides 14, 16, 18, and 22 are then incorporated into the synthesized fragment by DNA polymerase. Th long linker arms 14b,

16b, 18b, 22b isolate the fluorescent dye tags 14c, 16c, 18c, 22c from the bases 14a, 16a, 18a, 22a to permit uninhibited enzyme activity.

DNA fragments several kB long have been synthesized with each base containing a carbon chain linker arm terminating in biotin as hereinafter described. To exemplify the DNA synthesis, tagging, and cleaving processes a known strand of DNA nucleotides was formed, nucleotides were tagged with a linker arm terminating in biotin, and a complementary strand of DNA was synthesized from the tagged nucleotides. Biotin was used as a model tag rather than fluorescent dyes to demonstrate the synthesis and cleavage reactions.

1. Preparation of known strand [d(A,G)]:

A polydeoxynucleotide, $d(A,G)_{2138}$, was prepared by the method outlined in R. L. Ratliff et al., "Heteropolynucleotide Synthesis with Terminal Deoxyribonucleotidyltransferase," Biochemistry 6, 851 (1967) and "Heteropolynucleotides Synthesized with Terminal Deoxyribonucleotidyltransferase. II. Nearest Neighbor Frequencies and Extent of Digestion by Micrococcal Deoxyribonuclease," Biochemistry 7, 412 (1968). The subscript, 2138, refers to the average number of bases in the fragment and the comma between the A and the G indicates that the bases are incorporated in a random order.

Ten micromoles of the 5'-triphosphate of 2'-deoxyadenosine (dATP) were mixed with one micromole of the 5'-triphosphate of 2'-deoxyguanosine (dGTP) and 5.5 nanomoles of the linear heptamer of 5'-thymidylic acid $[d(pT)_7]$ which acts as a primer. Ten thousand units of terminal transferase were added to the solution which was buffered at pH 7 and the reaction mixture was maintained

at 37°C for 24 hours. (One unit is defined as the amount of enzyme which will polymerize 1 nanomole of nucleotide in one hour.) The resulting $d(A,G)_{2138}$ was then separated from the reaction mixture and purified.

2. Preparation of biotinylated complementary strand [$d(C,U)_{2138}$]:

The complementary strand of DNA to $d(A,G)_{2138}$, prepared as described above, was synthesized from nucleotides (dCTP) and d(UTP) tagged with biotin. A mixture of 10 nanomoles of the biotinylated 5'-triphosphate of 2'-deoxycytidine (dCTP) and 20 nanomoles of the biotinylated 5'-triphosphate of 2'-deoxyuridine (dUTP) was added to 10 nanomoles of $d(A,G)_{2138}$ and 22 picomoles of $d(pT)_7$. Ten units of DNA polymerase (E coli), Klenow fragment, were then added to the mixture which was buffered at pH 8 and maintained at a temperature of 37°C for 2 hours. Analysis of the resulting products by electrophoresis demonstrated that the reaction went to completion and the completely biotinylated complementary DNA fragment, $d(C,U)_{2138}$, was formed.

3. Exonuclease cleavage of biotinylated $d(C,U)_{2138}$:

The completely biotinylated $d(C,U)_{2138}$, synthesized as described above, was sequentially cleaved by adding 10 units of exonuclease III to 5 nanomoles of $d(A,G)_{2138}$ • biotinylated $d(C,U)_{2138}$. The reaction mixture was maintained at pH 8 and 37°C for two hours. At the end of two hours, analysis of the reaction mixture showed that 30% of the DNA was cleaved and the cleavage reaction appeared to be still proceeding. A control reaction using normal $d(C,T)_{2138}$ yielded 85% cleavage in two hours. Hence, biotinylation does appear

to slow the cleavage reaction using exonucl ase III, but the tagged nucleotides were sequentially cleaved from the DNA fragments.

In accordance with the present invention, the selected fluorescent dyes are substituted for biotin to specifically tag each nucleotide type with a dye characteristic of that nucleotide. The resulting complementary DNA chain will then provide each base with a characteristic, strongly fluorescing dye. By way of example, Smith et al., supra, teach a set of four individually distinguishable tags.

The sensitivity for fluorescence detection can be increased, if necessary, by attaching several dye molecules along the linker arm. Alternatively, large phycoerythrin-like molecules or even small microspheres containing many dye molecules may be attached to the linker arm. In yet another alternative, fluorescent labels might be attached to the primary, single stranded fragment, thereby eliminating the necessity of forming labeled bases and synthesizing the complementary strand.

It should be noted that DNA fragment 10 may be either a single or double strand of DNA. A single strand of DNA arises where the selected DNA strand is directly tagged for base identification or where the resynthesized complementary tagged DNA strand is separated from the selected strand. A double strand arises where the resynthesized DNA strand remains combined with the selected strand. As used herein, the term "fragment" refers to any and all of such conditions.

Enzymatic Cleavage of the Tagged Nucleotides

After DNA fragment 10 is formed with identifiable bases and hybridized to microsphere 40, a single fragment 10 can be manipulated with microsphere 40 and suspended in

flow stream 24. Exonuclease 20 is used to cleave bases 14a, 16a, 18a, 22a sequentially from single DNA fragment 10 suspended in flow stream 24. While the presence of the linker arm and the fluorescent dye may inhibit the enzymatic activity of some exonucleases, suitable exonucleases will cleave with only a slight reduction in rate. Individual bases have been sequentially enzymatically cleaved from DNA fragments formed completely from biotinylated nucleotides as demonstrated above. See, also, e.g., M. L. Shimkus et al., supra, incorporated herein by reference. The rate of cleavage can be adjusted by varying the exonuclease concentration, temperature, or by the use of poisoning agents. The time to remove one base can be made to be on the order of one millisecond. See, e.g., W. E. Razzell et al., "Studies on Polynucleotides," J. Bio. Chem. 234 No 8, 2105-2112 (1959).

Single Molecule Detection

The individual modified nucleotides 14, 16, 18, and 22 are carried by flow stream 24 into flow cell 26 for detection and analysis by single molecule detection system 34. One embodiment of a laser-induced fluorescence detection system is described in D. C. Nguyen et al., "Ultrasensitive Laser-Induced Fluorescence Detection in Hydrodynamically Focused Flows," J. Opt. Soc. Am. B, 4, 138-143, No. 2 (1987), incorporated herein by reference. The photomultiplier-based detection system described therein has detected single molecules of phycoerythrin in focused, flowing sample streams by laser-induced fluorescence. See D. C. Nguyen et al., "Detection of Single Molecules of Phycoerythrin in Hydrodynamically Focused Flows by Laser-Induced Fluorescence," Anal. Chem. 59, 2158-2161 (September 1987), incorporated herein by reference.

Phycoerythrin is a large protein containing the equivalent of 25 rhodamine-6G dye molecules. The detection of single molecules/chromophores of rhodamine-6G and equivalent dye molecules is suggested by system improvements. Thus, a combination of improved light collection efficiency, improved detector quantum efficiency, or pulsed excitation and gated detection to reduce background noise can be used with the Nguyen et al. system. Detection of phycoerythrin was accomplished in the 180 μs it took the molecule to flow through the focused laser beam.

In a preferred embodiment of the present process, the hydrodynamically focused flow system of Nguyen et al. is provided with an improved fluorescence detection system described in a copending patent application by Shera, "Single Molecule Tracking," Docket No. 65,737, incorporated herein by reference. As therein described, flow stream 24 provides to flow cell 26 modified nucleotides 14, 16, 18, and 22 in the sequence they are cleaved from DNA strand 10. Laser system 32 excites fluorescent dyes 14c, 16c, 18c, and 22c at selected wavelengths for identification in laminar sample flow 28 within flow cell 26.

Fluorescent events contained in optical signal 36 are focused by lens 38 on position sensitive detector system 42. Detector system 42 may comprise a microchannel plate (MCP) sensor to output spatial coordinates of observed photon events. An internal clock provides a temporal coordinate, wherein data processor 44 determines the presence of a molecule within flow cell 26. Molecular spectral response to laser 32 excitation enables the specific modified nucleotide to be identified. As noted by Shera, supra, data handling in the single molecule

detection system 34 effectively provides a moving sample volume within focused flow stream 28 which contains only a single tagged nucleotide. System 34 can thus track multiple molecules existing within focused flow stream 28 to enable a high rate of sequencing to be maintained.

Referring now to Figure 2, there is shown a representative output signal from the single molecule detection system. The individual nucleotide molecules 14, 16, 18, and 22 are individually cleaved from DNA strand 10 into flow stream 24. The flow velocity and laminar flow conditions maintain the molecules in a train for sequential passage through flow cell 26 and the emitted photons from laser-excited molecular fluorescence are assigned to individual molecules passing within the cell. The characteristic dye for each type nucleotide is selected to have an identifiable excitation or fluorescence spectrum. This characteristic spectrum can be used to establish the base sequence for the DNA strand being investigated.

It will be appreciated that the present process further provides a capability to sort the detected molecules and deposit them on a moving substrate for subsequent identification, e.g., as described in M. R. Melamed et al., "Flow Cytometry and Sorting," Wiley, New York (1979), incorporated herein by reference. The flow stream maintains the bases spatially isolated in a flow stream for presentation to a secondary identification device. The position between molecules on the moving substrate can be adjustable and can be large enough to resolve the sorted molecules by other techniques.

The foregoing description of the preferred embodiment of the inv ntion has been presented for purposes of illustration and description. It is not intended to be

exhaustive or to limit the invention to the precise form disclosed, and obviously many modifications and variations are possible in light of the above teaching. The embodiment was chosen and described in order to best explain the principles of the invention and its practical application to thereby enable others skilled in the art to best utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the claims appended hereto.

WHAT IS CLAIMED IS

1. A method for DNA and RNA base sequencing, comprising the steps of:

isolating a single fragment of DNA or RNA;

introducing said single fragment into a moving sample stream;

sequentially cleaving the end base from the DNA or RNA fragment with exonuclease to form a train of said bases; and

detecting said bases in said train in sequential passage through a detector.

2. A method according to Claim 1, wherein each said base of said single fragment is modified to contain a tag having an identifiable characteristic for said base.

3. A method according to Claim 2, where said bases are modified prior to said cleavage.

4. A method according to Claim 2, further including the step of enzymatically synthesizing a strand of DNA complementary to a DNA or RNA strand to be characterized, where each nucleotide forming said synthesized strand contains a tag characteristic of that nucleotide.

5. A method according to Claim 2, wherein said tag is separated from the nucleotide by a linker arm effective for said cleavage.

6. A method according to Claim 1, wherein said cleaved bases are detected optically.

7. A method according to Claim 6, wherein each said tag is a fluorescent dye characteristic of one type of said nucleotides.

8. A method according to Claim 7, further including the step of exciting each said fluorescent dye and detecting the fluorescence spectrum of said dye.

9. A method according to Claim 1, wherein said step of isolating said single fragment of DNA or RNA includes the step of hybridizing said fragment to a substrate having a site effective for said hybridization.

10. A method according to Claim 9, further including the step of selecting said DNA or RNA fragments from a heterogeneous collection of DNA or RNA fragments wherein said site is a biotinylated probe effective to hybridize with DNA or RNA fragments to be selected.

11. A method according to Claim 9, wherein said isolating said single fragment includes the step of providing said substrate with a single site effective to hybridize with a single DNA fragment.

12. A method for base sequencing of DNA or RNA fragments, comprising the steps of:

forming said fragments with bases having identifiable characteristics;

sequentially cleaving single identifiable bases from a single one of said fragments to form a train of said identifiable bases; and

identifying said single, cleaved bases in said train.

13. A method according to Claim 12, further including the step of attaching a characteristic identifiable fluorescent dye to each said base.

14. A method according to Claim 12, wherein the steps of forming said fragments includes the steps of forming by enzymatic synthesis a complementary strand of said DNA or RNA to be sequenced from said bases having identifiable characteristics and thereafter base sequencing said complementary strand.

15. A method according to Claim 14, further including the step of attaching a characteristic identifiable fluorescent dye to each said base.

16. A method according to Claim 13, wherein said step of identifying said single, cleaved bases includes the step of exciting each said fluorescent dye and detecting the fluorescence spectrum of said dye.

17. A method according to Claim 15, wherein said step of identifying said single, cleaved bases includes the step of exciting each said fluorescent dye and detecting the fluorescence spectrum of said dye.

18. A method for DNA or RNA base sequencing, comprising the steps of:

modifying each nucleotide effective for DNA or RNA synthesis to attach a fluorescent dye characteristic of that nucleotide with a linker arm effective to enable DNA or RNA synthesis and exonuclease cleavage;

synthesizing from said modified nucleotides a strand of DNA complementary to a DNA or RNA strand having a base sequence to be determined;

cleaving each said modified nucleotide sequentially from a single fragment containing said complementary DNA strand; and

fluorescing each said characteristic dye to identify said sequence of nucl otides.

19. A method according to Claim 18, wherein th step of fluorescing said dyes further comprises the steps of:

exciting each said modified nucleotide with a laser effective to fluoresce said characteristic dye; and

detecting said fluorescence to sequentially identify said nucleotides and generate said sequence of said DNA or RNA.

20. A method according to Claim 18, further including the step of suspending a single fragment of synthesized DNA or RNA strand in a laminar flow stream.

21. A method according to Claim 18, wherein each synthesized DNA or RNA fragment is manipulated by hybridizing said fragment to a microsphere having a site effective for hybridization.

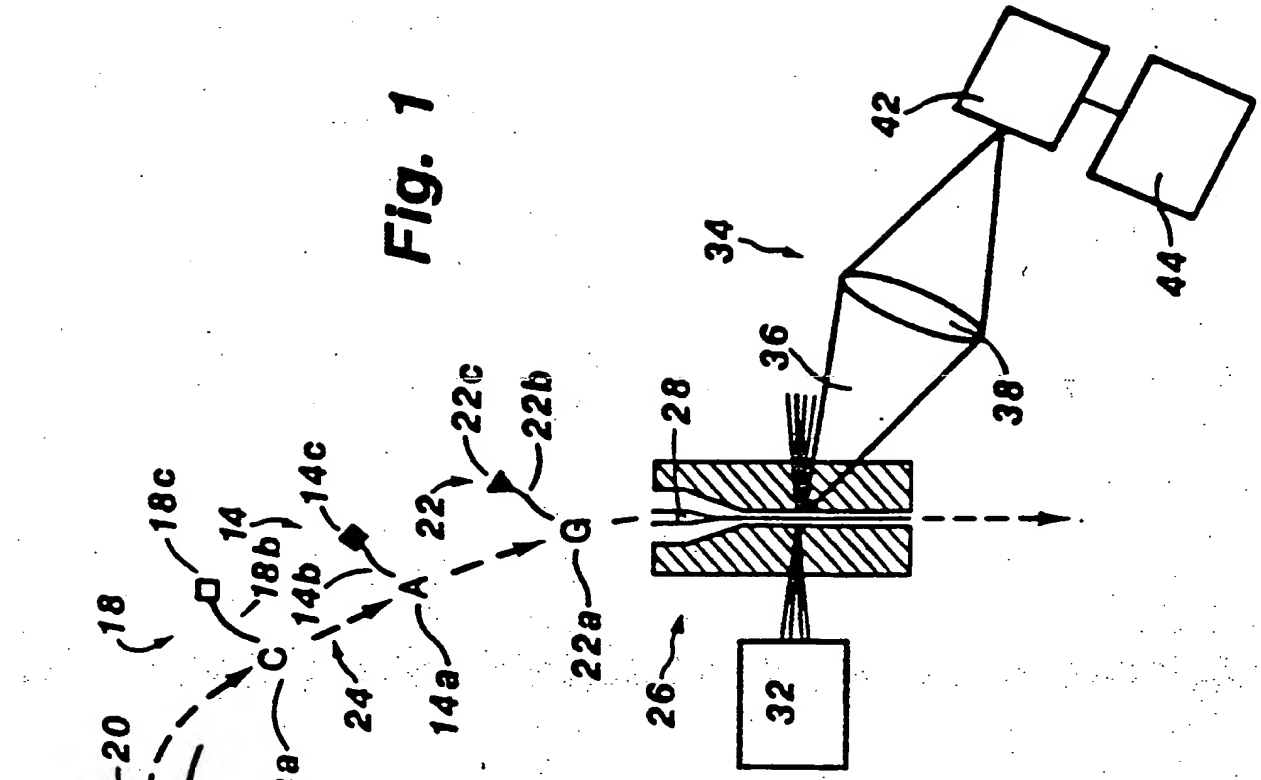

42
34
44
36
38
28
22c
22b
22
G
22a
26
32
18c
14
18b
14c
18
14b
A
24
14a
C
18a
20
16b
16c
16
16a
ATT
10a
14
12
10
40
9a
9

Fig. 1

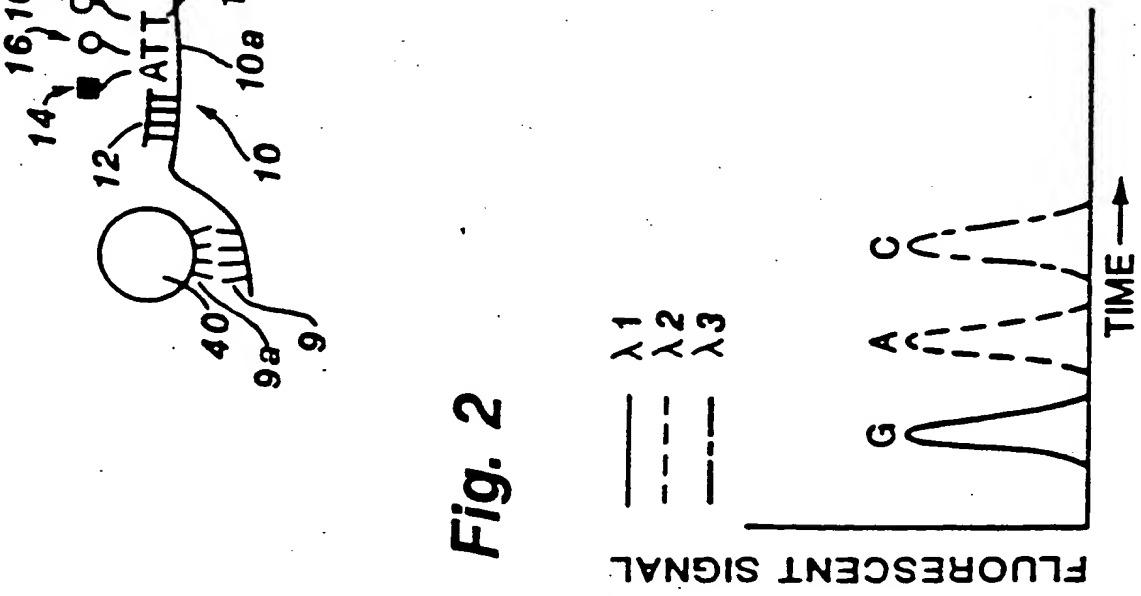

λ1
λ2
λ3
G
A
C
TIME
FLUORESCENT SIGNAL

Fig. 2